

Confidence Bands in Quantile Regression and Generalized Dynamic Semiparametric Factor Models

DISSERTATION

zur Erlangung des akademischen Grades

Doctorate

im Fach Statistics and Econometrics

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät

Humboldt-Universität zu Berlin

von

Msc. Song Song

18.07.1984

Präsident der Humboldt-Universität zu Berlin:

Prof. Dr. Dr. h.c. Christoph Marksches

Dekan der Wirtschaftswissenschaftlichen Fakultät:

Prof. Oliver Günther, Ph.D.

Gutachter:

1. Prof. Dr. Wolfgang Karl Härdle

2. Prof. Dr. Ya'acov Ritov

eingereicht am: 07.06.2010

Tag der mündlichen Prüfung: 02.08.2010

The topic this thesis deals with is complex, but very interesting. I want to thank my supervisor Prof. Dr. Wolfgang Karl Härdle and Prof. Dr. Ya'acov Ritov for the great opportunities to learn from them and the support during the production of this thesis. To overcome all of the theoretical and empirical challenges, I also want to thank Dr. T. Vogel, Peter Bickel for their great presence.

Abstract

In many applications it is necessary to know the stochastic fluctuation of the maximal deviations of the nonparametric quantile estimates, e.g. for various parametric models check. Uniform confidence bands are therefore constructed for nonparametric quantile estimates of regression functions. The first method is based on the strong approximations of the empirical process and extreme value theory. The strong uniform consistency rate is also established under general conditions. The second method is based on the bootstrap resampling method. It is proved that the bootstrap approximation provides a substantial improvement. The case of multidimensional and discrete regressor variables is dealt with using a partial linear model. A labor market analysis is provided to illustrate the method.

High dimensional time series which reveal nonstationary and possibly periodic behavior occur frequently in many fields of science, e.g. macroeconomics, meteorology, medicine and financial engineering. One of the common approach is to separate the modeling of high dimensional time series to time propagation of low dimensional time series and high dimensional time invariant functions via dynamic factor analysis. We propose a two-step estimation procedure. At the first step, we detrend the time series by incorporating time basis selected by the group Lasso-type technique and choose the space basis based on smoothed functional principal component analysis. We show properties of this estimator under the dependent scenario. At the second step, we obtain the detrended low dimensional stochastic process (stationary), but it also poses an important question: is it justified, from an inferential point of view, to base further statistical inference on the estimated stochastic time series? We show that the difference of the inference based on the estimated time series and “true” unobserved time series is asymptotically negligible, which finally allows one to study the dynamics of the whole high-dimensional system with a low dimensional representation together with the deterministic trend. We apply the method to our motivating empirical problems: studies of the dynamic behavior of temperatures (further used for pricing weather derivatives), implied volatilities and risk patterns and correlated brain activities (neuroeconomics related) using fMRI data, where a panel version model is also presented.

Zusammenfassung

In vielen Anwendungen ist es notwendig, die stochastische Schwankungen der maximalen Abweichungen der nichtparametrischen Schätzer von Quantil zu wissen, zB um die verschiedene parametrische Modelle zu überprüfen. Einheitliche Konfidenzbänder sind daher für nichtparametrische Quantil Schätzungen der Regressionsfunktionen gebaut. Die erste Methode basiert auf der starken Approximation der empirischen Verfahren und Extremwert-Theorie. Die starke gleichmäßige Konsistenz liegt auch unter allgemeinen Bedingungen etabliert. Die zweite Methode beruht auf der Bootstrap Resampling-Verfahren. Es ist bewiesen, dass die Bootstrap-Approximation eine wesentliche Verbesserung ergibt. Der Fall von mehrdimensionalen und diskrete Regressorvariablen wird mit Hilfe einer partiellen linearen Modell behandelt. Das Verfahren wird mithilfe der Arbeitsmarktanalysebeispiel erklärt.

Hoch-dimensionale Zeitreihen, die nichtstationäre und eventuell periodische Verhalten zeigen, sind häufig in vielen Bereichen der Wissenschaft, zB Makroökonomie, Meteorologie, Medizin und Financial Engineering, getroffen. Der typische Modellierungsansatz ist die Modellierung von hochdimensionalen Zeitreihen in Zeit Ausbreitung der niedrig dimensionalen Zeitreihen und hoch-dimensionale zeitinvarianten Funktionen über dynamische Faktorenanalyse zu teilen. Wir schlagen ein zweistufiges Schätzverfahren. Im ersten Schritt entfernen wir den Langzeittrend der Zeitreihen durch Einbeziehung Zeitbasis von der Gruppe Lasso-Technik und wählen den Raumbasis mithilfe der funktionalen Hauptkomponentenanalyse aus. Wir zeigen die Eigenschaften dieser Schätzer unter den abhängigen Szenario. Im zweiten Schritt erhalten wir den trendbereinigten niedrig-dimensionalen stochastischen Prozess (stationär). Allerdings bleibt auch eine wichtige Frage: ist es gerechtfertigt, von einer schließenden Sicht auf weiteren statistischen Rückschlüssen auf der geschätzten stochastischen Zeitreihen zu stützen? Wir zeigen, dass die Differenz des Schlusses basiert auf den geschätzten Zeitreihen und den "wahren" unbeobachteten Zeitreihen sind asymptotisch vernachlässigbar, die schließlich erlaubt uns, die Dynamik des gesamten hochdimensionalen Systems mit einer niedrig-dimensionalen Darstellung zusammen mit dem deterministischen Trend zu studieren. Wir wenden die Methode, zu empirischen Problemen wie zB Untersuchungen des dynamischen Verhaltens von Temperaturen (weitere Preise für Wetterderivate verwendet), die implizite Volatilität und Risiko Muster und korreliert Hirnaktivitäten (Neuroökonomie verwandt) mit Hilfe von fMRI-Daten, wo ein Panel Version Modell ist auch vorgelegt, an.

Contents

1	Confidence Bands in Quantile Regression	1
1.1	Introduction	1
1.2	Results	3
1.3	A Monte Carlo Study	9
1.4	Application	10
1.5	Appendix	11
2	Partial Linear Quantile Regression and Bootstrap Confidence Bands	19
2.1	Introduction	19
2.2	Bootstrap confidence bands in the univariate case	21
2.3	Bootstrap confidence bands in PLMs	25
2.4	A Monte Carlo study	28
2.5	A labour market application	33
2.6	Appendix	39
3	Generalized Dynamic Semiparametric Factor Models	45
3.1	Introduction	45
3.2	Methodology	50
3.2.1	Choice of Time Basis	50
3.2.2	Choice of Space Basis	50
3.2.3	Estimation Procedure	51
3.3	Estimates' Properties	53
3.4	Generalized Dynamic Semiparametric Factor Model	57
3.5	Simulation Study	59
3.6	Weather, Neuro-economics and IVS	61
3.7	Appendix	69

1 Confidence Bands in Quantile Regression

1.1 Introduction

In standard regression function estimation, most investigations are concerned with the conditional mean regression. However, new insights about the underlying structures can be gained by considering other aspects of the conditional distribution. The quantile curves are key aspects of inference in various economic problems and are of great interest in practice. These describe the conditional behavior of a response variable (e.g. wage of workers) given the value of an explanatory variable (e.g. education level, experience, occupation of workers), and investigate changes in both tails of the distribution, other than just the mean.

When examining labour markets, economists are concerned with whether discrimination exists, for example for different genders, nationalities, union status and so on. To study this question, we need to separate out other effects first, e.g. age, education, etc. The crucial relation between age and earnings or salaries belongs to the most carefully studied subjects in labor economics. The fundamental work in mean regression can be found in Murphy and Welch (1990). Quantile regression estimates could provide more accurate measures. Koenker and Hallock (2001) present a basket of important economic applications, including quantile Engel curves and claim that “quantile regression is gradually developing into a comprehensive strategy for completing the regression prediction”. Besides this, it is also well-known that a quantile regression model (e.g. the conditional median curve) is more robust to outliers, especially for fat-tailed distributions. For symmetric conditional distributions the quantile regression generates the nonparametric mean regression analysis since the $p = 0.5$ (median) quantile curve coincides with the mean regression.

As first introduced by Koenker and Bassett (1978), one may assume a parametric model for the p -quantile curve and estimate parameters by the interior point method discussed by Koenker and Park (1996) and Portnoy and Koenker (1997). Similarly, we can also adopt nonparametric methods to estimate conditional quantiles. The first one, a more direct approach using a “check” function such as a robustified local linear smoother, is provided by Fan et al. (1994) and further extended by Yu and Jones (1997, 1998). An alternative procedure is first to estimate the conditional distribution function using the “double-kernel” local linear technique of Fan et al. (1996) and then to invert the conditional distribution estimator to produce an estimator of a conditional quantile by Yu and Jones (1997, 1998). Beside these, Hall et al. (1999) proposed a weighted version of the Nadaraya-Watson estimator, which was further studied by Cai (2002). Recently Jeong et al. (2009) have developed the conditional quantile causality test. More generally, for an M -regression function which involves quantile regression as a

special case, the uniform Bahadur representation and application to the additive model is studied by Kong et al. (2008). An interesting question for the parametric fitting, especially from labour economists, would be how well these models fit the data, when compared with the nonparametric estimation method.

Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a sequence of independent identically distributed bivariate random variables with joint pdf $f(x, y)$, joint cdf $F(x, y)$, conditional pdf $f(y|x)$, $f(x|y)$, conditional cdf $F(y|x)$, $F(x|y)$ for Y given X and X given Y respectively, and marginal pdf $f_X(x)$ for X , $f_Y(y)$ for Y where $x \in J$, and J is a possibly infinite interval in \mathbb{R}^d and $y \in \mathbb{R}$. In general, X may be a multivariate covariate, although here we restrict attention to the univariate case and $J = [0, 1]$ for convenience. Let $l(x)$ denote the p -quantile curve, i.e. $l(x) = F_{Y|x}^{-1}(p)$.

Under a “check function”, the quantile regression curve $l(x)$ can be viewed as the minimiser of $L(\theta) \stackrel{\text{def}}{=} \mathbb{E}\{\rho_p(y - \theta)|X = x\}$ (w.r.t. θ) with $\rho_p(u) = pu\mathbf{1}\{u \in (0, \infty)\} - (1 - p)u\mathbf{1}\{u \in (-\infty, 0)\}$ which was originally motivated by an exercise in Ferguson (1967)[p.51] in the literature.

A kernel-based p -quantile curve estimator $l_n(x)$ can naturally be constructed by minimising:

$$L_n(\theta) = n^{-1} \sum_{i=1}^n \rho_p(Y_i - \theta) K_h(x - X_i) \quad (1.1)$$

with respect to $\theta \in I$ where I is a possibly infinite, or possibly degenerate, interval in \mathbb{R} , and $K_h(u) = h^{-1}K(u/h)$ is a kernel with bandwidth h . The numerical solution of (1.1) may be found iteratively as in Lejeune and Sarda (1988) and Yu et al. (2003).

In light of the concepts of M -estimation as in Huber (1981), if we define $\psi(u)$ as:

$$\begin{aligned} \psi_p(u) &= p\mathbf{1}\{u \in (0, \infty)\} - (1 - p)\mathbf{1}\{u \in (-\infty, 0)\} \\ &= p - \mathbf{1}\{u \in (-\infty, 0)\}, \end{aligned}$$

$l_n(x)$ and $l(x)$ can be treated as a zero (w.r.t. θ) of the function:

$$\tilde{H}_n(\theta, x) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n K_h(x - X_i) \psi(Y_i - \theta) \quad (1.2)$$

$$\tilde{H}(\theta, x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(x, y) \psi(y - \theta) dy \quad (1.3)$$

correspondingly.

To show the uniform consistency of the quantile smoother, we shall reduce the problem of strong convergence of $l_n(x) - l(x)$, uniformly in x , to an application of the strong convergence of $\tilde{H}_n(\theta, x)$ to $\tilde{H}(\theta, x)$, uniformly in x and θ , as given by Theorem 2.2 in Härdle et al. (1988). It is shown that under general conditions almost surely (a.s.)

$$\sup_{x \in J} |l_n(x) - l(x)| \leq B^* \max\{(nh/(\log n))^{-1/2}, h^{\tilde{\alpha}}\}, \quad \text{as } n \rightarrow \infty.$$

where B^* and $\tilde{\alpha}$ are parameters defined more precisely in Subsection 1.2.

Please note that without assuming K has compact support (as we do here) under similar assumptions Franke and Mwita (2003) get:

$$\begin{aligned} l_n(x) &= \hat{F}_{Y|x}^{-1}(p) \\ \hat{F}(y|x) &= \frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}(Y_i < y)}{\sum_{i=1}^n K_h(x - X_i)} \\ \sup_{x \in J} |l_n(x) - l(x)| &\leq B^{**} \{(nh/(s_n \log n))^{-1/2} + h^2\}, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

for α -mixing data where B^{**} is some constant and $s_n, n \geq 1$ is an increasing sequence of positive integers satisfying $1 \leq s_n \leq \frac{n}{2}$ and some other criteria. Thus $\{nh/(\log n)\}^{-1/2} \leq \{nh/(s_n \log n)\}^{-1/2}$.

By employing similar methods as those developed in Härdle (1989) it is shown in this thesis that

$$\begin{aligned} &P\left((2\delta \log n)^{1/2} \left[\sup_{x \in J} r(x) |\{l_n(x) - l(x)\}| / \lambda(K)^{1/2} - d_n \right] < z\right) \\ &\rightarrow \exp\{-2 \exp(-z)\}, \quad \text{as } n \rightarrow \infty. \end{aligned} \tag{1.4}$$

from the asymptotic Gumbel distribution where $r(x)$, δ , $\lambda(K)$, d_n are suitable scaling parameters. The asymptotic result (1.4) therefore allows the construction of (asymptotic) uniform confidence bands for $l(x)$ based on specifications of the stochastic fluctuation of $l_n(x)$. The strong approximation with Brownian bridge techniques that we use in this thesis is available only for the approximation of the 2-dimensional empirical process. The extension to the multivariate covariable can be done by partial linear modelling which deserves further research.

The plan of the thesis is as follows. In Subsection 1.2, the stochastic fluctuation of the process $\{l_n(x) - l(x)\}$ and the uniform confidence band are presented through the equivalence of several stochastic processes, with a strong uniform consistency rate of $\{l_n(x) - l(x)\}$ also shown. In Subsection 1.3, in a small Monte Carlo study we investigate the behaviour of $l_n(x)$ when the data is generated by fat-tailed conditional distributions of $(Y|X = x)$. In Subsection 1.4, an application considers a wage-earning relation in the labour market. All proofs are sketched in Subsection 1.5.

1.2 Results

The following assumptions will be convenient. To make x and X clearly distinguishable, we replace x by t sometimes, but they are essentially the same.

- (A1) The kernel $K(\cdot)$ is positive, symmetric, has compact support $[-A, A]$ and is Lipschitz continuously differentiable with bounded derivatives;
- (A2) $(nh)^{-1/2}(\log n)^{3/2} \rightarrow 0$, $(n \log n)^{1/2} h^{5/2} \rightarrow 0$, $(nh^3)^{-1}(\log n)^2 \leq M$, M a constant;
- (A3) $h^{-3}(\log n) \int_{|y| > a_n} f_Y(y) dy = \mathcal{O}(1)$, $f_Y(y)$ the marginal density of Y , $\{a_n\}_{n=1}^\infty$ a sequence of constants tending to infinity as $n \rightarrow \infty$;

1 Confidence Bands in Quantile Regression

(A4) $\inf_{t \in J} |q(t)| \geq q_0 > 0$, where $q(t) = \partial \mathbb{E}\{\psi(Y - \theta)|t\} / \partial \theta|_{\theta=l(t)} \cdot f_X(t)$
 $= f\{l(t)|t\} f_X(t)$;

(A5) the quantile function $l(t)$ is Lipschitz twice continuously differentiable, for all $t \in J$.

(A6) $0 < m_1 \leq f_X(t) \leq M_1 < \infty$, $t \in J$; the conditional densities $f(\cdot|y)$, $y \in \mathbb{R}$, are uniform local Lipschitz continuous of order $\tilde{\alpha}$ (ulL- $\tilde{\alpha}$) on J , uniformly in $y \in \mathbb{R}$, with $0 < \tilde{\alpha} \leq 1$.

Define also

$$\begin{aligned}\sigma^2(t) &= \mathbb{E}[\psi^2\{Y - l(t)\}|t] = p(1 - p) \\ H_n(t) &= (nh)^{-1} \sum_{i=1}^n K\{(t - X_i)/h\} \psi\{Y_i - l(t)\} \\ D_n(t) &= \partial(nh)^{-1} \sum_{i=1}^n K\{(t - X_i)/h\} \psi\{Y_i - \theta\} / \partial \theta|_{\theta=l(t)}\end{aligned}$$

and assume that $\sigma^2(t)$ and $f_X(t)$ are differentiable.

Assumption (A1) on the compact support of the kernel could possibly be relaxed by introducing a cutoff technique as in Csörgö and Hall (1982) for density estimators. Assumption (A2) has purely technical reasons: to keep the bias at a lower rate than the variance and to ensure the vanishing of some non-linear remainder terms. Assumption (A3) appears in a somewhat modified form also in Johnston (1982). Assumptions (A5, A6) are common assumptions in robust estimation as in Huber (1981), Härdle et al. (1988) that are satisfied by exponential, and generalised hyperbolic distributions.

For the uniform strong consistency rate of $l_n(x) - l(x)$, we apply the result of Härdle et al. (1988) by taking $\beta(y) = \psi(y - \theta)$, $y \in \mathbb{R}$, for $\theta \in I = \mathbb{R}$, $q_1 = q_2 = -1$, $\gamma_1(y) = \max\{0, -\psi(y - \theta)\}$, $\gamma_2(y) = \min\{0, -\psi(y - \theta)\}$ and $\lambda = \infty$ to satisfy the representations for the parameters there. Thus from Theorem 2.2 and Remark 2.3(v) there we immediately have the following lemma.

LEMMA 1.2.1 *Let $\tilde{H}_n(\theta, x)$ and $\tilde{H}(\theta, x)$ be given by (1.2) and (1.3). Under assumption (A6) and $(nh/\log n)^{-1/2} \rightarrow \infty$ through (A2), for some constant A^* not depending on n , we have a.s. as $n \rightarrow \infty$*

$$\sup_{\theta \in I} \sup_{x \in J} |\tilde{H}_n(\theta, x) - \tilde{H}(\theta, x)| \leq A^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \quad (1.5)$$

For our result on $l_n(\cdot)$, we shall also require

$$\inf_{x \in J} \left| \int \psi\{y - l(x) + \varepsilon\} dF(y|x) \right| \geq \tilde{q}|\varepsilon|, \quad \text{for } |\varepsilon| \leq \delta_1, \quad (1.6)$$

where δ_1 and \tilde{q} are some positive constants, see also Härdle and Luckhaus (1984). This assumption is satisfied if there exists a constant \tilde{q} such that $f(l(x)|x) > \tilde{q}/p$, $x \in J$.

THEOREM 1.2.1 *Under the conditions of Lemma 1.2.1 and also assuming (1.6), we have a.s. as $n \rightarrow \infty$*

$$\sup_{x \in J} |l_n(x) - l(x)| \leq B^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \quad (1.7)$$

with $B^* = A^*/m_1\tilde{q}$ not depending on n and m_1 a lower bound of $f_X(t)$. If additionally $\tilde{\alpha} \geq \{\log(\sqrt{\log n}) - \log(\sqrt{nh})\}/\log h$, it can be further simplified to

$$\sup_{x \in J} |l_n(x) - l(x)| \leq B^* \{(nh/\log n)^{-1/2}\}.$$

THEOREM 1.2.2 *Let $h = n^{-\delta}$, $\frac{1}{5} < \delta < \frac{1}{3}$, $\lambda(K) = \int_{-A}^A K^2(u)du$ and*

$$\begin{aligned} d_n &= (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} [\log\{c_1(K)/\pi^{1/2}\} + \frac{1}{2}\{\log \delta + \log \log n\}], \\ &\text{if } c_1(K) = \{K^2(A) + K^2(-A)\}/\{2\lambda(K)\} > 0 \\ d_n &= (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \log\{c_2(K)/2\pi\} \\ &\text{otherwise with } c_2(K) = \int_{-A}^A \{K'(u)\}^2 du / \{2\lambda(K)\}. \end{aligned}$$

Then (1.4) holds with

$$r(x) = (nh)^{1/2} f\{l(x)|x\} \{f_X(x)/p(1-p)\}^{1/2}.$$

This theorem can be used to construct uniform confidence intervals for the regression function as stated in the following corollary.

COROLLARY 1.2.1 *Under the assumptions of the theorem above, an approximate $(1 - \alpha) \times 100\%$ confidence band over $[0, 1]$ is*

$$l_n(t) \pm (nh)^{-1/2} \{p(1-p)\lambda(K)/\hat{f}_X(t)\}^{1/2} \hat{f}^{-1}\{l(t)|t\} \{d_n + c(\alpha)(2\delta \log n)^{-1/2}\},$$

where $c(\alpha) = \log 2 - \log |\log(1 - \alpha)|$ and $\hat{f}_X(t)$, $\hat{f}\{l(t)|t\}$ are consistent estimates for $f_X(t)$, $f\{l(t)|t\}$.

In the literature, according to Fan et al. (1994, 1996), Yu and Jones (1997, 1998), Hall et al. (1999), Cai (2002) and others, asymptotic normality at interior points for various nonparametric smoothers, e.g. local constant, local linear, reweighted NW methods, etc. has been shown:

$$\sqrt{nh}\{l_n(t) - l(t)\} \sim N(0, \tau^2(t))$$

with $\tau^2(t) = \lambda(K)p(1-p)/[f_X(t)f^2\{l(t)|t\}]$. Please note that the bias term vanishes here as we adjust h . With $\tau(t)$ introduced, we can further write Corollary 1.2.1 as:

$$l_n(t) \pm (nh)^{-1/2} \{d_n + c(\alpha)(2\delta \log n)^{-1/2}\} \hat{\tau}(t).$$

1 Confidence Bands in Quantile Regression

Through minimising the approximation of AMSE (asymptotic mean square error), the optimal bandwidth h_p can be computed. In practice, the rule-of-thumb for h_p is given by Yu and Jones (1998):

1. Use ready-made and sophisticated methods to select optimal bandwidth h_{mean} from conditional mean regression, e.g. Ruppert et al. (1995)
2. $h_p = [p(1-p)/\varphi^2\{\Phi^{-1}(p)\}]^{1/5} \cdot h_{\text{mean}}$
with φ, Φ as the pdf and cdf of a standard normal distribution

Obviously the further p lies from 0.5, the more smoothing is necessary.

The proof is essentially based on a linearisation argument after a Taylor series expansion. The leading linear term will then be approximated in a similar way as in Johnston (1982), Bickel and Rosenblatt (1973). The main idea behind the proof is a strong approximation of the empirical process of $\{(X_i, Y_i)_{i=1}^n\}$ by a sequence of Brownian bridges as proved by Tusnady (1977).

As $l_n(t)$ is the zero (w.r.t. θ) of $\tilde{H}_n(\theta, t)$, it follows by applying 2nd-order Taylor expansions to $\tilde{H}_n(\theta, t)$ around $l(t)$ that

$$l_n(t) - l(t) = \{H_n(t) - \mathbf{E} H_n(t)\}/q(t) + R_n(t) \quad (1.8)$$

where $\{H_n(t) - \mathbf{E} H_n(t)\}/q(t)$ is the leading linear term and

$$R_n(t) = H_n(t)\{q(t) - D_n(t)\}/\{D_n(t) \cdot q(t)\} + \mathbf{E} H_n(t)/q(t) + \frac{1}{2}\{l_n(t) - l(t)\}^2 \cdot \{D_n(t)\}^{-1} \quad (1.9)$$

$$\cdot (nh)^{-1} \sum_{i=1}^n K\{(x - X_i)/h\} \psi''\{Y_i - l(t) + r_n(t)\}, \quad (1.10)$$

$$|r_n(t)| < |l_n(t) - l(t)|.$$

is the remainder term. In Subsection 1.5 it is shown (Lemma 1.5.1) that $\|R_n\| = \sup_{t \in J} |R_n(t)| = \mathcal{O}_p\{(nh \log n)^{-1/2}\}$.

Furthermore, the rescaled linear part

$$Y_n(t) = (nh)^{1/2} \{\sigma^2(t) f_X(t)\}^{-1/2} \{H_n(t) - \mathbf{E} H_n(t)\}$$

is approximated by a sequence of Gaussian processes, leading finally to the Gaussian process

$$Y_{5,n}(t) = h^{-1/2} \int K\{(t - x)/h\} dW(x). \quad (1.11)$$

Drawing upon the result of Bickel and Rosenblatt (1973), we finally obtain asymptotically the Gumbel distribution.

We also need the Rosenblatt (1952) transformation,

$$T(x, y) = \{F_{X|y}(x|y), F_Y(y)\},$$

which transforms (X_i, Y_i) into $T(X_i, Y_i) = (X'_i, Y'_i)$ mutually independent uniform rv's. In the event that x is a d -dimension covariate, the transformation becomes:

$$T(x_1, x_2, \dots, x_d, y) = \{F_{X_1|y}(x_1|y), F_{X_2|y}(x_2|x_1, y), \dots, F_{X_k|x_{d-1}, \dots, x_1, y}(x_k|x_{d-1}, \dots, x_1, y), F_Y(y)\}. \quad (1.12)$$

With the aid of this transformation, Theorem 1 of Tusnady (1977) may be applied to obtain the following lemma.

LEMMA 1.2.2 *On a suitable probability space a sequence of Brownian bridges B_n exists that*

$$\sup_{x \in J, y \in \mathbb{R}} |Z_n(x, y) - B_n\{T(x, y)\}| = \mathcal{O}\{n^{-1/2}(\log n)^2\} \quad a.s.,$$

where $Z_n(x, y) = n^{1/2}\{F_n(x, y) - F(x, y)\}$ denotes the empirical process of $\{(X_i, Y_i)\}_{i=1}^n$.

For $d > 2$, it is still an open problem which deserves further research.

Before we define the different approximating processes, let us first rewrite (1.11) as a stochastic integral w.r.t. the empirical process $Z_n(x, y)$,

$$Y_n(t) = \{hg'(t)\}^{-1/2} \iint K\{(t-x)/h\}\psi\{y-l(t)\}dZ_n(x, y),$$

$$g'(t) = \sigma^2(t)f_X(t).$$

1 Confidence Bands in Quantile Regression

The approximating processes are now:

$$Y_{0,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\} \psi\{y-l(t)\} dZ_n(x, y) \quad (1.13)$$

$$\text{where } \Gamma_n = \{|y| \leq a_n\}, g(t) = E[\psi^2\{y-l(t)\} \cdot \mathbf{1}(|y| \leq a_n) | X=t] \cdot f_X(t)$$

$$Y_{1,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\} \psi\{y-l(t)\} dB_n\{T(x, y)\} \quad (1.14)$$

$\{B_n\}$ being the sequence of Brownian bridges from Lemma 1.2.2.

$$Y_{2,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\} \psi\{y-l(t)\} dW_n\{T(x, y)\} \quad (1.15)$$

$\{W_n\}$ being the sequence of Wiener processes satisfying

$$B_n(x', y') = W_n(x', y') - x'y'W_n(1, 1)$$

$$Y_{3,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\} \psi\{y-l(x)\} dW_n\{T(x, y)\} \quad (1.16)$$

$$Y_{4,n}(t) = \{hg(t)\}^{-1/2} \int g(x)^{1/2} K\{(t-x)/h\} dW(x) \quad (1.17)$$

$$Y_{5,n}(t) = h^{-1/2} \int K\{(t-x)/h\} dW(x) \quad (1.18)$$

$\{W(\cdot)\}$ being the Wiener process.

Lemmas 1.5.2 to 1.5.7 ensure that all these processes have the same limit distributions. The result then follows from

LEMMA 1.2.3 (*Theorem 3.1 in Bickel and Rosenblatt (1973)*) *Let d_n , $\lambda(K)$, δ as in Theorem 1.2.2. Let*

$$Y_{5,n}(t) = h^{-1/2} \int K\{(t-x)/h\} dW(x).$$

Then, as $n \rightarrow \infty$, the supremum of $Y_{5,n}(t)$ has a Gumbel distribution.

$$P \left\{ (2\delta \log n)^{1/2} \left[\sup_{t \in J} |Y_{5,n}(t)| / \{\lambda(K)\}^{1/2} - d_n \right] < z \right\} \rightarrow \exp\{-2 \exp(-z)\}.$$

1.3 A Monte Carlo Study

We generate bivariate data $\{(X_i, Y_i)\}_{i=1}^n, n = 500$ with joint pdf:

$$\begin{aligned} f(x, y) &= g(y - \sqrt{x + 2.5}) \mathbf{1}(x \in [-2.5, 2.5]) \\ g(u) &= \frac{9}{10} \varphi(u) + \frac{1}{90} \varphi(u/9). \end{aligned} \quad (1.19)$$

The p -quantile curve $l(x)$ can be obtained from a zero (w.r.t. θ) of:

$$9\Phi(\theta) + \Phi(\theta/9) = 10p,$$

with Φ as the cdf of a standard normal distribution. Solving it numerically gives the 0.5-quantile curve $l(x) = \sqrt{x + 2.5}$, and the 0.9-quantile curve $l(x) = 1.5296 + \sqrt{x + 2.5}$. We use the quartic kernel:

$$\begin{aligned} K(u) &= \frac{15}{16}(1 - u^2)^2, \quad |u| \leq 1, \\ &= 0, \quad |u| > 1. \end{aligned}$$

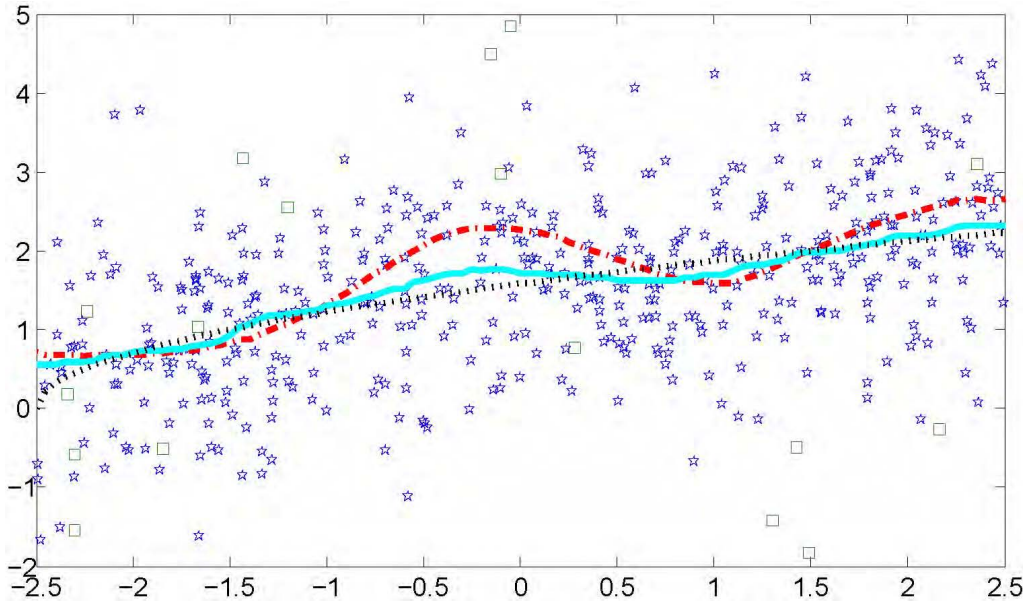


Figure 1.1: The 0.5-quantile curve, the Nadaraya-Watson estimator $m_n^*(x)$, and the 0.5-quantile smoother $l_n(x)$.

In Fig. 1.1 the raw data, together with the 0.5-quantile curve, are displayed. The random variables generated with probability $\frac{1}{10}$ from the fat-tailed pdf $\frac{1}{9}\varphi(u/9)$, see (1.19), are marked as squares whereas the standard normal rv's are shown as stars. We then compute both the Nadaraya-Watson estimator $m_n^*(x)$ and the 0.5-quantile smoother

$l_n(x)$. The bandwidth is set to 1.25 which is equivalent to 0.25 after rescaling x to $[0, 1]$ and fulfills the requirements of Theorem 1.2.2.

In Fig. 1.1 $l(x)$, $m_n^*(x)$ and $l_n(x)$ are shown as a dotted line, dashed-dot line, and solid line respectively. At first sight $m_n^*(x)$ has clearly more variation and has the expected sensitivity to the fat-tails of $f(x, y)$. A closer look reveals that $m_n^*(x)$ for $x \approx 0$ apparently even leaves the 0.5-quantile curve. It may be surprising that this happens at $x \approx 0$ where no outlier is placed, but a closer look at Fig. 1.1 shows that the large negative data values at both $x \approx -0.1$ and $x \approx 0.25$ cause the problem. This data value is inside the window ($h = 1.10$) and therefore distorts $m_n^*(x)$ for $x \approx 0$. The quantile-smoother $l_n(x)$ (solid line) is unaffected and stays fairly close to the 0.5-quantile curve. Similar results can be obtained in Fig. 1.2 corresponding to the 0.9 quantile ($h = 1.25$) with the 95% confidence band.

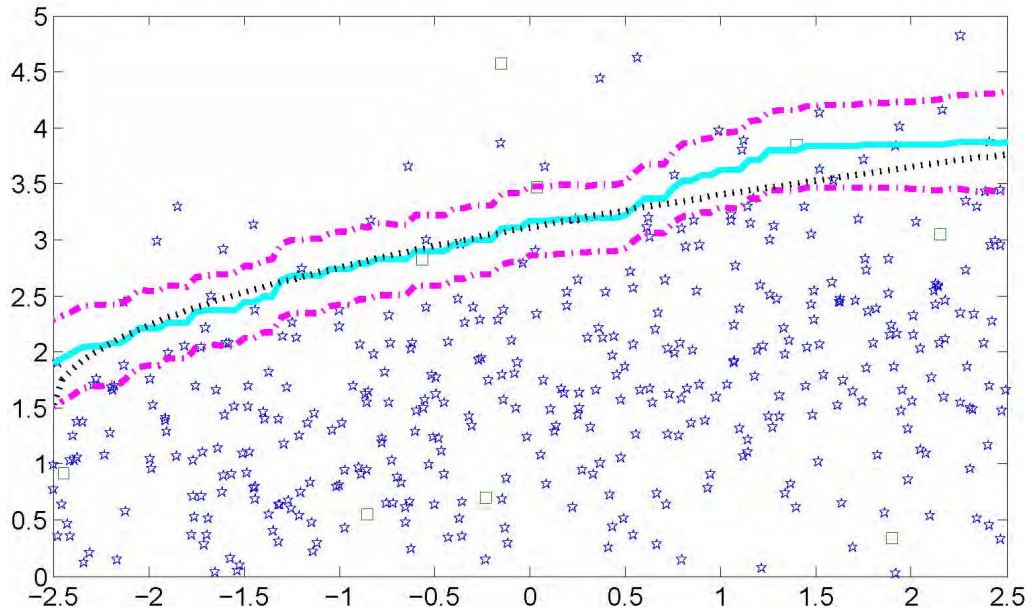


Figure 1.2: The 0.9-quantile curve, the 0.9-quantile smoother and 95% confidence band.

1.4 Application

Recently there has been great interest in finding out how the financial returns of a job depend on the age of the employee. We use the Current Population Survey (CPS) data from 2005 for the following group: male aged 25 – 59, full-time employed, and college graduate containing 16,731 observations, for the age-earning estimation. As is usual for wage data, a log transformation to hourly real wages (unit: US dollar) is carried out first. In the CPS all ages (25 ~ 59) are reported as integers. We rescaled them into $[0, 1]$ by dividing 40 by bandwidth 0.059 for nonparametric quantile-smoothers. This is equivalent to set bandwidth 2 for the original age data.

In Fig. 1.3 the original observations are displayed as small stars. The local 0.5 and 0.9 quantiles at the integer points of age are shown as dashed lines, whereas the corresponding nonparametric quantile-smoothers are displayed as solid lines with corresponding 95% uniform confidence bands shown as dashed-dot lines. A closer look reveals a quadratic relation between age and logged hourly real wages. If we use several popular parametric methods to estimate the 0.5 and 0.9 conditional quantiles, e.g. quadratic, quartic and set of dummies (a dummy variable for each 5-year age group) models as in Fig. 1.4. With the help of the 95% uniform confidence bands, we can do the parametric model specification test. At the 5% significance level, we could not reject any model. However, when the confidence level further decreases and the uniform confidence bands get narrower, “set of dummies” parametric model will be the first one to be rejected. At the 10% significance level, the set of dummies (for age groups) model is rejected while the other two are not. As the quadratic model performs quite similar by the quartic one, for simplicity, it is suggested in practice to measure the log(wage)-earring relation which coincides with Murphy and Welch (1990) in mean regression.

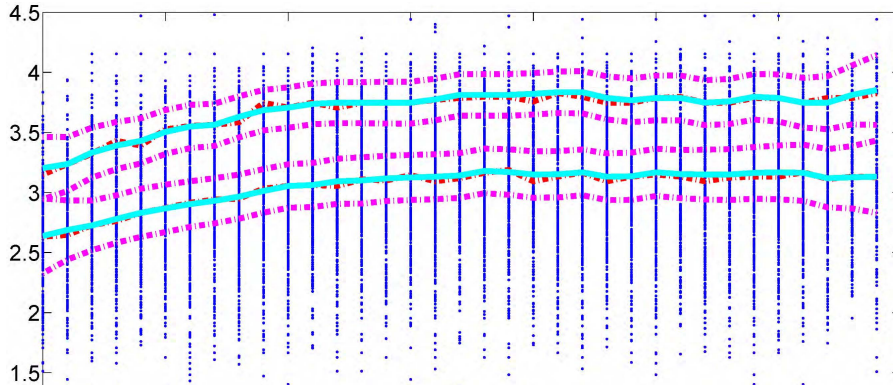


Figure 1.3: The original observations, local quantiles, 0.5, 0.9-quantile smoothers and corresponding 95% confidence bands.

1.5 Appendix

Proof of Theorem 1.2.1. By the definition of $l_n(x)$ as a zero of (1.2), we have, for $\varepsilon > 0$,

$$\text{if } l_n(x) > l(x) + \varepsilon, \text{ and then } \tilde{H}_n\{l(x) + \varepsilon, x\} > 0. \quad (1.20)$$

Now

$$\tilde{H}_n\{l(x) + \varepsilon, x\} \leq \tilde{H}\{l(x) + \varepsilon, x\} + \sup_{\theta \in I} |\tilde{H}_n(\theta, x) - \tilde{H}(\theta, x)|. \quad (1.21)$$

Also, by the identity $\tilde{H}\{l(x), x\} = 0$, the function $\tilde{H}\{l(x) + \varepsilon, x\}$ is not positive and has a magnitude $\geq m_1 \tilde{q} \varepsilon$ by assumption (A6) and (1.6), for $0 < \varepsilon < \delta_1$. That is, for $0 < \varepsilon < \delta_1$,

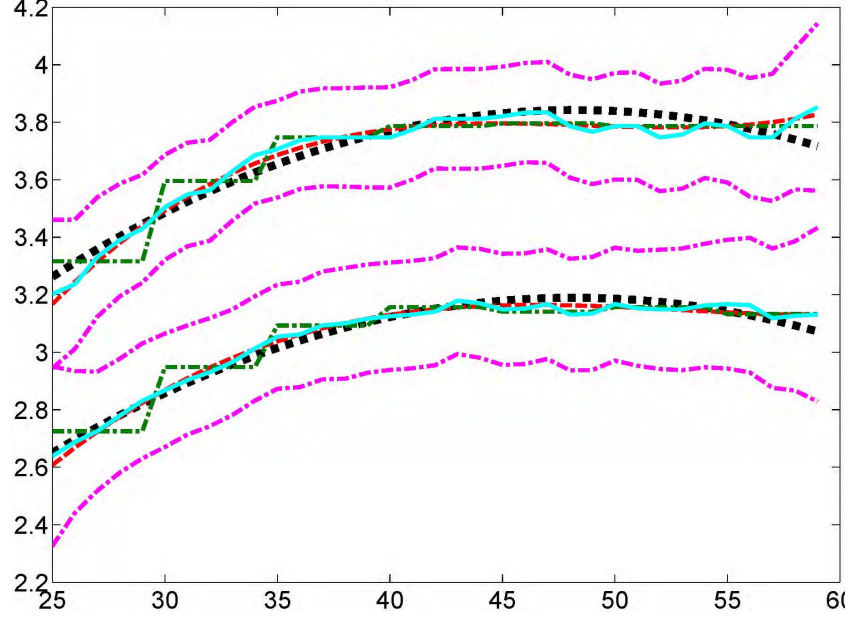


Figure 1.4: Quadratic, quartic, set of dummies (for age groups) estimates, 0.5, 0.9-quantile smoothers and their corresponding 95% confidence bands.

$$\tilde{H}\{l(x) + \varepsilon, x\} \leq -m_1 \tilde{q} \varepsilon. \quad (1.22)$$

Combining (1.20), (1.21) and (1.22), we have, for $0 < \varepsilon < \delta_1$:

$$\text{if } l_n(x) > l(x) + \varepsilon, \text{ and then } \sup_{\theta \in I} \sup_{x \in J} |\tilde{H}_n(\theta, x) - \tilde{H}(\theta, x)| > m_1 \tilde{q} \varepsilon.$$

With a similar inequality proved for the case $l_n(x) < l(x) + \varepsilon$, we obtain, for $0 < \varepsilon < \delta_1$:

$$\text{if } \sup_{x \in J} |l_n(x) - l(x)| > \varepsilon, \text{ and then } \sup_{\theta \in I} \sup_{x \in J} |\tilde{H}_n(\theta, x) - \tilde{H}(\theta, x)| > m_1 \tilde{q} \varepsilon. \quad (1.23)$$

It readily follows that (1.23), and (1.5) imply (1.7). \blacksquare

Below we first show that $\|R_n\|_\infty = \sup_{t \in J} |R_n(t)|$ vanishes asymptotically faster than the rate $(nh \log n)^{-1/2}$; for simplicity we will just use $\|\cdot\|$ to indicate the sup-norm.

LEMMA 1.5.1 *For the remainder term $R_n(t)$ defined in (1.9) we have*

$$\|R_n\| = o_p\{(nh \log n)^{-1/2}\}. \quad (1.24)$$

Proof First we have by the positivity of the kernel K ,

$$\begin{aligned} \|R_n\| &\leq \left[\inf_{0 \leq t \leq 1} \{|D_n(t)| \cdot q(t)\} \right]^{-1} \{ \|H_n\| \cdot \|q - D_n\| + \|D_n\| \cdot \| \mathbf{E} H_n \| \} \\ &\quad + C_1 \cdot \|l_n - l\|^2 \cdot \left\{ \inf_{0 \leq t \leq 1} |D_n(t)| \right\}^{-1} \cdot \|f_n\|_\infty, \end{aligned}$$

where $f_n(x) = (nh)^{-1} \sum_{i=1}^n K\{(x - X_i)/h\}$.

The desired result (1.5.1) will then follow if we prove

$$\|H_n\| = \mathcal{O}_p\{(nh)^{-1/2}(\log n)^{1/2}\} \quad (1.25)$$

$$\|q - D_n\| = \mathcal{O}_p\{(nh)^{-1/4}(\log n)^{-1/2}\} \quad (1.26)$$

$$\| \mathbf{E} H_n \| = \mathcal{O}(h^2) \quad (1.27)$$

$$\|l_n - l\|^2 = \mathcal{O}_p\{(nh)^{-1/2}(\log n)^{-1/2}\} \quad (1.28)$$

Since (1.27) follows from the well-known bias calculation

$$\mathbf{E} H_n(t) = h^{-1} \int K\{(t - u)/h\} \mathbf{E}[\psi\{y - l(t)\}|X = u] f_X(u) du = \mathcal{O}(h^2),$$

where $\mathcal{O}(h^2)$ is independent of t in Parzen (1962), we have from assumption (A2) that $\| \mathbf{E} H_n \| = \mathcal{O}_p\{(nh)^{-1/2}(\log n)^{-1/2}\}$.

According to Lemma A.3 in Franke and Mwita (2003),

$$\sup_{t \in J} |H_n(t) - \mathbf{E} H_n(t)| = \mathcal{O}\{(nh)^{-1/2}(\log n)^{1/2}\}.$$

and the following inequality

$$\begin{aligned} \|H_n\| &\leq \|H_n - \mathbf{E} H_n\| + \| \mathbf{E} H_n \|. \\ &= \mathcal{O}\{(nh)^{-1/2}(\log n)^{1/2}\} + \mathcal{O}_p\{(nh)^{-1/2}(\log n)^{-1/2}\} \\ &= \mathcal{O}\{(nh)^{-1/2}(\log n)^{1/2}\} \end{aligned}$$

Statement (1.25) thus is obtained.

Statement (1.26) follows in the same way as (1.25) using assumption (A2) and the Lipschitz continuity properties of K , ψ' , l .

According to the uniform consistency of $l_n(t) - l(t)$ shown before, we have

$$\|l_n - l\| = \mathcal{O}_p\{(nh)^{-1/2}(\log n)^{1/2}\}$$

which implies (1.28).

Now the assertion of the lemma follows, since by tightness of $D_n(t)$, $\inf_{0 \leq t \leq 1} |D_n(t)| \geq q_0$ *a.s.* and thus

$$\|R_n\| = \mathcal{O}_p\{(nh \log n)^{-1/2}\}(1 + \|f_n\|).$$

Finally, by Theorem 3.1 of Bickel and Rosenblatt (1973), $\|f_n\| = \mathcal{O}_p(1)$; thus the desired result $\|R_n\| = \mathcal{O}_p\{(nh \log n)^{-1/2}\}$ follows. \square

We now begin with the subsequent approximations of the processes $Y_{0,n}$ to $Y_{5,n}$.

LEMMA 1.5.2

$$\|Y_{0,n} - Y_{1,n}\| = \mathcal{O}\{(nh)^{-1/2}(\log n)^2\} \quad a.s.$$

Proof Let t be fixed and put $L(y) = \psi\{y - l(t)\}$ still depending on t . Using integration by parts, we obtain

$$\begin{aligned} & \iint_{\Gamma_n} L(y) K\{(t-x)/h\} dZ_n(x, y) \\ &= \int_{u=-A}^A \int_{y=-a_n}^{a_n} L(y) K(u) dZ_n(t - h \cdot u, y) \\ &= - \int_{-A}^A \int_{-a_n}^{a_n} Z_n(t - h \cdot u, y) d\{L(y) K(u)\} \\ &+ L(a_n)(a_n) \int_{-A}^A Z_n(t - h \cdot u, a_n) dK(u) \\ &- L(-a_n)(-a_n) \int_{-A}^A Z_n(t - h \cdot u, -a_n) dK(u) \\ &+ K(A) \left\{ \int_{-a_n}^{a_n} Z_n(t - h \cdot A, y) dL(y) \right. \\ &+ L(a_n)(a_n) Z_n(t - h \cdot A, a_n) - L(-a_n)(-a_n) Z_n(t - h \cdot A, -a_n) \left. \right\} \\ &- K(-A) \left\{ \int_{-a_n}^{a_n} Z_n(t + h \cdot A, y) dL(y) + L(a_n)(a_n) Z_n(t + h \cdot A, a_n) \right. \\ &- L(-a_n)(-a_n) Z_n(t + h \cdot A, -a_n) \left. \right\}. \end{aligned}$$

If we apply the same operation to $Y_{1,n}$ with $B_n\{T(x, y)\}$ instead of $Z_n(x, y)$ and use Lemma 1.2.2, we finally obtain

$$\sup_{0 \leq t \leq 1} h^{1/2} g(t)^{1/2} |Y_{0,n}(t) - Y_{1,n}(t)| = \mathcal{O}\{n^{-1/2}(\log n)^2\} \quad a.s..$$

\square

LEMMA 1.5.3 $\|Y_{1,n} - Y_{2,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof Note that the Jacobian of $T(x, y)$ is $f(x, y)$. Hence

$$\begin{aligned} & Y_{1,n}(t) - Y_{2,n}(t) \\ &= \left| \{g(t)h\}^{-1/2} \iint_{\Gamma_n} \psi\{y - l(t)\} K\{(t-x)/h\} f(x, y) dx dy \right| \cdot |W_n(1, 1)|. \end{aligned}$$

It follows that

$$h^{-1/2}\|Y_{1,n} - Y_{2,n}\| \leq |W_n(1, 1)| \cdot \|g^{-1/2}\| \cdot \sup_{0 \leq t \leq 1} h^{-1} \iint_{\Gamma_n} |\psi\{y - l(t)\} K\{(t - x)/h\}| f(x, y) dx dy.$$

Since $\|g^{-1/2}\|$ is bounded by assumption, we have

$$h^{-1/2}\|Y_{1,n} - Y_{2,n}\| \leq |W_n(1, 1)| \cdot C_4 \cdot h^{-1} \int K\{(t - x)/h\} dx = \mathcal{O}_p(1).$$

□

LEMMA 1.5.4 $\|Y_{2,n} - Y_{3,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof The difference $|Y_{2,n}(t) - Y_{3,n}(t)|$ may be written as

$$\left| \{g(t)h\}^{-1/2} \iint_{\Gamma_n} [\psi\{y - l(t)\} - \psi\{y - l(x)\}] K\{(t - x)/h\} dW_n\{T(x, y)\} \right|.$$

If we use the fact that l is uniformly continuous, this is smaller than

$$h^{-1/2}|g(t)|^{-1/2} \cdot \mathcal{O}_p(h)$$

and the lemma thus follows. □

LEMMA 1.5.5 $\|Y_{4,n} - Y_{5,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof

$$\begin{aligned} |Y_{4,n}(t) - Y_{5,n}(t)| &= h^{-1/2} \left| \int \left[\left\{ \frac{g(x)}{g(t)} \right\}^{1/2} - 1 \right] K\{(t - x)/h\} dW(x) \right| \\ &\leq h^{-1/2} \left| \int_{-A}^A W(t - hu) \frac{\partial}{\partial u} \left[\left\{ \frac{g(t - hu)}{g(t)} \right\}^{1/2} - 1 \right] K(u) du \right| \\ &\quad + h^{-1/2} \left| K(A)W(t - hA) \left[\left\{ \frac{g(t - Ah)}{g(t)} \right\}^{1/2} - 1 \right] \right| \\ &\quad + h^{-1/2} \left| K(-A)W(t + hA) \left[\left\{ \frac{g(t + Ah)}{g(t)} \right\}^{1/2} - 1 \right] \right| \\ &= S_{1,n}(t) + S_{2,n}(t) + S_{3,n}(t), \quad \text{say.} \end{aligned}$$

The second term can be estimated by

$$h^{-1/2}\|S_{2,n}\| \leq K(A) \cdot \sup_{0 \leq t \leq 1} |W(t - Ah)| \cdot \sup_{0 \leq t \leq 1} h^{-1} \left| \left[\left\{ \frac{g(t - Ah)}{g(t)} \right\}^{1/2} - 1 \right] \right|;$$

1 Confidence Bands in Quantile Regression

by the mean value theorem it follows that

$$h^{-1/2}\|S_{2,n}\| = \mathcal{O}_p(1).$$

The first term $S_{1,n}$ is estimated as

$$\begin{aligned} h^{-1/2}S_{1,n}(t) &= \left| h^{-1} \int_{-A}^A W(t-uh)K'(u) \left[\left\{ \frac{g(t-uh)}{g(t)} \right\}^{1/2} - 1 \right] du \right. \\ &\quad \left. \frac{1}{2} \int_{-A}^A W(t-uh)K(u) \left\{ \frac{g(t-uh)}{g(t)} \right\}^{1/2} \left\{ \frac{g'(t-uh)}{g(t)} \right\} du \right| \\ &= |T_{1,n}(t) - T_{2,n}(t)|, \quad \text{say}; \end{aligned}$$

$\|T_{2,n}\| \leq C_5 \cdot \int_{-A}^A |W(t-hu)|du = \mathcal{O}_p(1)$ by assumption on $g(t) = \sigma^2(t) \cdot f_X(t)$. To estimate $T_{1,n}$ we again use the mean value theorem to conclude that

$$\sup_{0 \leq t \leq 1} h^{-1} \left| \left\{ \frac{g(t-uh)}{g(t)} \right\}^{1/2} - 1 \right| < C_6 \cdot |u|;$$

hence

$$\|T_{1,n}\| \leq C_6 \cdot \sup_{0 \leq t \leq 1} \int_{-A}^A |W(t-hu)|K'(u)u/du = \mathcal{O}_p(1).$$

Since $S_{3,n}(t)$ is estimated as $S_{2,n}(t)$, we finally obtain the desired result. \square

The next lemma shows that the truncation introduced through $\{a_n\}$ does not affect the limiting distribution.

LEMMA 1.5.6 $\|Y_n - Y_{0,n}\| = \mathcal{O}_p\{(\log n)^{-1/2}\}.$

Proof We shall only show that $g'(t)^{-1/2}h^{-1/2} \iint_{\mathbb{R}-\Gamma_n} \psi\{y-l(t)\}K\{(t-x)/h\}dZ_n(x,y)$ fulfills the lemma. The replacement of $g'(t)$ by $g(t)$ may be proved as in Lemma A.4 of Johnston (1982). The quantity above is less than $h^{-1/2}\|g^{-1/2}\| \cdot \|\iint_{\{|y|>a_n\}} \psi\{y-l(\cdot)\}K\{(\cdot-x)/h\}dZ(x,y)\|$. It remains to be shown that the last factor tends to zero at a rate $\mathcal{O}_p\{(\log n)^{-1/2}\}$. We show first that

$$\begin{aligned} V_n(t) &= (\log n)^{1/2}h^{-1/2} \iint_{\{|y|>a_n\}} \psi\{y-l(t)\}K\{(t-x)/h\}dZ_n(x,y) \\ &\xrightarrow{p} 0 \quad \text{for all } t \end{aligned}$$

and then we show tightness of $V_n(t)$, the result then follows:

$$\begin{aligned} V_n(t) &= (\log n)^{1/2} (nh)^{-1/2} \sum_{i=1}^n [\psi\{Y_i - l(t)\} \mathbf{1}(|Y_i| > a_n) K\{(t - X_i)/h\} \\ &\quad - \mathbb{E} \psi\{Y_i - l(t)\} \mathbf{1}(|Y_i| > a_n) K\{(t - X_i)/h\}] \\ &= \sum_{i=1}^n X_{n,t}(t), \end{aligned}$$

where $\{X_{n,t}(t)\}_{i=1}^n$ are i.i.d. for each n with $\mathbb{E} X_{n,t}(t) = 0$ for all $t \in [0, 1]$. We then have

$$\begin{aligned} \mathbb{E} X_{n,t}^2(t) &\leq (\log n)(nh)^{-1} \mathbb{E} \psi^2\{Y_i - l(t)\} \mathbf{1}(|Y_i| > a_n) K^2\{(t - X_i)/h\} \\ &\leq \sup_{-A \leq u \leq A} K^2(u) \cdot (\log n)(nh)^{-1} \mathbb{E} \psi^2\{Y_i - l(t)\} \mathbf{1}(|Y_i| > a_n); \end{aligned}$$

hence

$$\begin{aligned} \text{Var}\{V_n(t)\} &= \mathbb{E} \left\{ \sum_{i=1}^n X_{n,t}(t) \right\}^2 = n \cdot \mathbb{E} X_{n,t}^2(t) \\ &\leq \sup_{-A \leq u \leq A} K^2(u) h^{-1} (\log n) \int_{\{|y| > a_n\}} f_y(y) dy \cdot M_\psi. \end{aligned}$$

where M_ψ denotes an upper bound for ψ^2 . This term tends to zero by assumption (A3). Thus by Markov's inequality we conclude that

$$V_n(t) \xrightarrow{P} 0 \quad \text{for all } t \in [0, 1].$$

To prove tightness of $\{V_n(t)\}$ we refer again to the following moment condition as stated in Lemma 1.5.1:

$$\begin{aligned} \mathbb{E}\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\} &\leq C' \cdot (t_2 - t_1)^2 \\ C' \text{ denoting a constant, } t &\in [t_1, t_2]. \end{aligned}$$

We again estimate the left-hand side by Schwarz's inequality and estimate each factor separately,

$$\begin{aligned} \mathbb{E}\{V_n(t) - V_n(t_1)\}^2 &= (\log n)(nh)^{-1} \mathbb{E} \left[\sum_{i=1}^n \Psi_n(t, t_1, X_i, Y_i) \cdot \mathbf{1}(|Y_i| > a_n) \right. \\ &\quad \left. - \mathbb{E}\{\Psi_n(t, t_1, X_i, Y_i) \cdot \mathbf{1}(|Y_i| > a_n)\} \right]^2, \end{aligned}$$

where $\Psi_n(t, t_1, X_i, Y_i) = \psi\{Y_i - l(t)\} K\{(t - X_i)/h\} - \psi\{Y_i - l(t_1)\} K\{(t_1 - X_i)/h\}$. Since ψ, K are Lipschitz continuous except at one point and the expectation is taken

1 Confidence Bands in Quantile Regression

afterwards, it follows that

$$\begin{aligned} & [\mathbb{E}\{V_n(t) - V_n(t_1)\}^2]^{1/2} \\ & \leq C_7 \cdot (\log n)^{1/2} h^{-3/2} |t - t_1| \cdot \left\{ \int_{\{|y| > a_n\}} f_y(y) dy \right\}^{1/2}. \end{aligned}$$

If we apply the same estimation to $V_n(t_2) - V_n(t_1)$ we finally have

$$\begin{aligned} & \mathbb{E}\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\} \\ & \leq C_7^2 (\log n) h^{-3} |t - t_1| |t_2 - t| \times \int_{\{|y| > a_n\}} f_y(y) dy \\ & \leq C' \cdot |t_2 - t_1|^2 \text{ since } t \in [t_1, t_2] \quad \text{by (A3)}. \end{aligned}$$

□

LEMMA 1.5.7 *Let $\lambda(K) = \int K^2(u) du$ and let $\{d_n\}$ be as in the theorem. Then*

$$(2\delta \log n)^{1/2} [\|Y_{3,n}\| / \{\lambda(K)\}^{1/2} - d_n]$$

has the same asymptotic distribution as

$$(2\delta \log n)^{1/2} [\|Y_{4,n}\| / \{\lambda(K)\}^{1/2} - d_n].$$

Proof $Y_{3,n}(t)$ is a Gaussian process with

$$\mathbb{E} Y_{3,n}(t) = 0$$

and covariance function

$$\begin{aligned} r_3(t_1, t_2) &= \mathbb{E} Y_{3,n}(t_1) Y_{3,n}(t_2) \\ &= \{g(t_1)g(t_2)\}^{-1/2} h^{-1} \iint_{\Gamma_n} \psi^2\{y - l(x)\} K\{(t_1 - x)/h\} \\ &\quad \times K\{(t_2 - x)/h\} f(x, y) dx dy \\ &= \{g(t_1)g(t_2)\}^{-1/2} h^{-1} \iint_{\Gamma_n} \psi^2\{y - l(x)\} f(y|x) dy K\{(t_1 - x)/h\} \\ &\quad \times K\{(t_2 - x)/h\} f_X(x) dx \\ &= \{g(t_1)g(t_2)\}^{-1/2} h^{-1} \int g(x) K\{(t_1 - x)/h\} K\{(t_2 - x)/h\} dx \\ &= r_4(t_1, t_2) \end{aligned}$$

where $r_4(t_1, t_2)$ is the covariance function of the Gaussian process $Y_{4,n}(t)$, which proves the lemma. □

2 Partial Linear Quantile Regression and Bootstrap Confidence Bands

2.1 Introduction

Quantile regression, as first introduced by Koenker and Bassett (1978), is “gradually developing into a comprehensive strategy for completing the regression prediction” as claimed by Koenker and Hallock (2001). Quantile smoothing is an effective method to estimate quantile curves in a flexible nonparametric way. Since this technique makes no structural assumptions on the underlying curve, it is very important to have a device for understanding when observed features are significant and deciding between functional forms, for example a question often asked in this context is whether or not an observed peak or valley is actually a feature of the underlying regression function or is only an artifact of the observational noise. For such issues, confidence intervals should be used that are simultaneous (i.e., uniform over location) in nature. Moreover, uniform confidence bands give an idea about the global variability of the estimate.

In the previous work the theoretical focus has mainly been on obtaining consistency and asymptotic normality of the quantile smoother, thereby providing the necessary ingredients to construct its pointwise confidence intervals. This, however, is not sufficient to get an idea about the global variability of the estimate, neither can it be used to correctly answer questions about the curve’s shape, which contains the lack of fit test as an immediate application. This motivates us to construct the confidence bands. To this end, Härdle and Song (2010) used strong approximations of the empirical process and extreme value theory. However, the very poor convergence rate of extremes of a sequence of n independent normal random variables is well documented and was first noticed and investigated by Fisher and Tippett (1928), and discussed in greater detail by Hall (1991). In the latter thesis it was shown that the rate of the convergence to its limit (the suprema of a stationary Gaussian process) can be no faster than $(\log n)^{-1}$. For example, the supremum of a nonparametric quantile estimate can converge to its limit no faster than $(\log n)^{-1}$. These results may make extreme value approximation of the distributions of suprema somewhat doubtful, for example in the context of the uniform confidence band construction for a nonparametric quantile estimate.

This thesis proposes and analyzes a method of obtaining any number of uniform confidence bands for quantile estimates. The method is simple to implement, does not rely on the evaluation of quantities which appear in asymptotic distributions and also takes the bias properly into account (at least asymptotically). More importantly, we show that the bootstrap approximation to the distribution of the supremum of a quantile estimate is accurate to within $n^{-2/5}$ which represents a significant improvement relative

to $(\log n)^{-1}$. Previous research by Hahn (1995) showed consistency of a bootstrap approximation to the cumulative density function (cdf) without assuming independence of the error and regressor terms. Horowitz (1998) showed bootstrap methods for median regression models based on a smoothed least-absolute-deviations (SLAD) estimate.

Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a sequence of independent identically distributed bivariate random variables with joint pdf $f(x, y)$, joint cdf $F(x, y)$, conditional pdf $f(y|x)$, $f(x|y)$, conditional cdf $F(y|x)$, $F(x|y)$ for Y given X and X given Y respectively, and marginal pdf $f_X(x)$ for X , $f_Y(y)$ for Y . With some abuse of notation we use the letters f and F to denote different pdf's and cdf's respectively. The exact distribution will be clear from the context. At the first stage we assume that $x \in J^*$, and $J^* = (a, b)$ for some $0 < a < b < 1$. Let $l(x)$ denote the p -quantile curve, i.e. $l(x) = F_{Y|x}^{-1}(p)$.

In economics, discrete or categorical regressors are very common. An example is from labour market analyse where one tries to find out how revenues depend on the age of the employee (for different education levels, labour union status, genders and nationalities), i.e. in econometric analysis one targets for the differential effects. For example, Buchinsky (1995) examined the U.S. wage structure by quantile regression techniques. This motivates the extension to multivariate covariables by partial linear modelling (PLM). This is convenient especially when we have categorical elements of the X vector. Partial linear models, which were first considered by Green and Yandell (1985), Denby (1986), Speckman (1988) and Robinson (1988), are gradually developing into a class of commonly used and studied semiparametric regression models, which can retain the flexibility of nonparametric models and ease the interpretation of linear regression models while avoiding the ‘‘curse of dimensionality’’. Recently Liang and Li (2009) used penalised quantile regression for variable selection of partially linear models with measurement errors.

In this thesis, we propose an extension of the quantile regression model to $x = (u, v)^\top \in \mathbb{R}^d$ with $u \in \mathbb{R}^{d-1}$ and $v \in J^* \subset \mathbb{R}$. The quantile regression curve we consider is: $\tilde{l}(x) = F_{Y|x}^{-1}(p) = u^\top \beta + l(v)$. The multivariate confidence band can now be constructed, based on the univariate uniform confidence band, plus the estimated linear part which we will prove is more accurately (\sqrt{n} consistency) estimated. This makes various tasks in economics, e.g. labour market differential effect investigation, multivariate model specification tests and the investigation of the distribution of income and wealth across regions, countries or the distribution across households possible. Additionally, since the natural link between quantile and expectile regression was developed by Newey and Powell (1987), we can further extend our result into expectile regression for various tasks, e.g. demography risk research or expectile-based Value at Risk (EVAR) as in Kuan et al. (2009). For high-dimensional modelling, Belloni and Chernozhukov (2009) recently investigated high-dimensional sparse models with L_1 penalty (LASSO). Additionally, by simple calculations, our result can be further extended to intersection bounds (one side confidence bands), which is similar to Chernozhukov et al. (2009).

The rest of this article is organised as follows. To keep the main idea transparent, we start with Subsection 2.2, as an introduction to the more complicated situation, the bootstrap approximation rate for the uniform confidence band (univariate case) in quantile regression is presented through a coupling argument. An extension to multivariate

covariance X with partial linear modelling is shown in Subsection 2.3 with the actual type of confidence bands and their properties. In Subsection 2.4, in the Monte Carlo study we compare the bootstrap uniform confidence band with the one based on the asymptotic theory and investigate the behaviour of partial linear estimates with the corresponding confidence band. In Subsection 2.5, an application considers the labour market differential effect. The discussion is restricted to the semiparametric extension. We do not discuss the general nonparametric regression. We conjecture that this extension is possible under appropriate conditions. All proofs are sketched in Subsection 2.6.

2.2 Bootstrap confidence bands in the univariate case

Suppose $Y_i = l(X_i) + \varepsilon_i$, $i = 1, \dots, n$, where ε_i has distribution function $F(\cdot|X_i)$. For simplicity, but without any loss of generality, we assume that $F(0|X_i) = p$. $F(\xi|x)$ is smooth as a function of x and ξ for any x , and for any ξ in the neighbourhood of 0. We assume:

1. X_1, \dots, X_n are an i.i.d. sample, and $\inf_x f_X(x) = \lambda_0 > 0$. The quantile function satisfies: $\sup_x |l^{(j)}(x)| \leq \lambda_j < \infty$, $j = 1, 2$.
2. The distribution of Y given X has a density and $\inf_{x,t} f(t|x) \geq \lambda_3 > 0$, continuous at all $x \in J^*$, and at t only in a neighbourhood of 0. More exactly, we have the following Taylor expansion, for some $A(\cdot)$ and $f_0(\cdot)$, and for every x, x', t :

$$F(t|x') = p + f_0(x)t + A(x)(x' - x) + R(t, x'; x), \quad (2.1)$$

where

$$\sup_{t,x,x'} \frac{|R(t, x'; x)|}{t^2 + |x' - x|^2} < \infty.$$

Let K be a symmetric density function with compact support and $d_K = \int u^2 K(u) du < \infty$. Let $l_h(\cdot) = l_{n,h}(\cdot)$ be the nonparametric p -quantile estimate of Y_1, \dots, Y_n with weight function $K\{(X_i - \cdot)/h\}$ for some global bandwidth $h = h_n$ ($K_h(u) = h^{-1}K(u/h)$), that is, a solution of:

$$\frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}\{Y_i < l_h(x)\}}{\sum_{i=1}^n K_h(x - X_i)} < p \leq \frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}\{Y_i \leq l_h(x)\}}{\sum_{i=1}^n K_h(x - X_i)}. \quad (2.2)$$

Generally, the bandwidth may also depend on x . A local (adaptive) bandwidth selection though deserves future research.

Note that by assumption (A1), $l_h(x)$ is the quantile of a discrete distribution, which is equivalent to a sample of size $\mathcal{O}_p(nh)$ from a distribution with p -quantile whose bias is $\mathcal{O}(h^2)$ relative to the true value. Let δ_n be the local rate of convergence of the function l_h , essentially $\delta_n = h^2 + (nh)^{-1/2} = \mathcal{O}(n^{-2/5})$ with optimal bandwidth choice $h = h_n = \mathcal{O}(n^{-1/5})$. We employ also an auxiliary estimate $l_g \stackrel{\text{def}}{=} l_{n,g}$, essentially one similar to

$l_{n,h}$ but with a slightly larger bandwidth $g = g_n = h_n n^\zeta$ (a heuristic explanation of why it is essential to oversmooth g is given later), where ζ is some small number. The asymptotically optimal choice of ζ as shown later is $4/45$.

3. The estimate l_g satisfies:

$$\begin{aligned} \sup_{x \in J^*} |l_g''(x) - l''(x)| &= \mathcal{O}_p(1), \\ \sup_{x \in J^*} |l_g'(x) - l'(x)| &= \mathcal{O}_p(\delta_n/h). \end{aligned} \quad (2.3)$$

Assumption (**A3**) is only stated to overwrite the issue here. It actually follows from the assumptions on (g, h) . A sequence $\{a_n\}$ is slowly varying if $n^{-\alpha} a_n \rightarrow 0$ for any $\alpha > 0$. With some abuse of notation we will use S_n to denote *any* slowly varying function which may change from place to place e.g. $S_n^2 = S_n$ is a valid expression (since if S_n is a slowly varying function, then S_n^2 is slowly varying as well). λ_i and C_i are generic constants throughout this thesis and the subscripts have no specific meaning. Note that there is no S_n term in (2.3) exactly because the bandwidth g_n used to calculate l_g is slightly larger than that used for l_h . We want to smooth it such that, l_g , as an estimate of the quantile function, has a slightly worse rate of convergence, but its derivatives converge faster.

We also consider a family of estimates $\hat{F}(\cdot|X_i)$, $i = 1, \dots, n$, estimating respectively $F(\cdot|X_i)$ and satisfying $\hat{F}(0|X_i) = p$. For example we can take the distribution with a point mass $c^{-1} K\{\alpha_n(X_j - X_i)\}$ on $Y_j - l_h(X_i)$, $j = 1, \dots, n$, where $c = \sum_{j=1}^n K\{\alpha_n(X_j - X_i)\}$ and $\alpha_n \approx h^{-1}$, i.e.

$$\hat{F}(\cdot|X_i) = \frac{\sum_{j=1}^n K_h(X_j - X_i) \mathbf{1}\{Y_j - l_h(X_i) \leq \cdot\}}{\sum_{j=1}^n K_h(X_j - X_i)}$$

We additionally assume:

4. $f_X(x)$ is twice continuously differentiable and $f(t|x)$ is uniformly bounded in x and t by, say, λ_4 .

LEMMA 2.2.1 [Franke and Mwita (2003), p14] *If assumptions (A1, A2, A4) hold, then for any small enough (positive) $\varepsilon \rightarrow 0$,*

$$\sup_{|t| < \varepsilon, i=1, \dots, n, X_i \in J^*} |\hat{F}(t|X_i) - F(t|X_i)| = \mathcal{O}_p\{S_n \delta_n \varepsilon^{1/2} + \varepsilon^2\}. \quad (2.4)$$

Note that the result in Lemma 2.2.1 is natural, since by definition, there is no error at $t = 0$, since $\hat{F}(0|X_i) \equiv p \equiv F(0|X_i)$. For $t \in (0, \varepsilon)$, $\hat{F}(t|X_i)$, like l_h , is based on a sample of size $\mathcal{O}_p(nh)$. Hence, the random error is $\mathcal{O}_p\{(nh)^{-1/2} t^{1/2}\}$, while the bias is $\mathcal{O}_p(\varepsilon h^2) = \mathcal{O}_p(\delta_n)$. The S_n term takes care of the maximisation.

Let $F^{-1}(\cdot|\cdot)$ and $\hat{F}^{-1}(\cdot|\cdot)$ be the inverse function of the conditional cdf and its estimate. We consider the following bootstrap procedure: Let U_1, \dots, U_n be i.i.d. uniform $[0, 1]$ variables. Let

$$Y_i^* = l_g(X_i) + \hat{F}^{-1}(U_i|X_i), \quad i = 1, \dots, n \quad (2.5)$$

2.2 Bootstrap confidence bands in the univariate case

be the bootstrap sample. We couple this sample to an unobserved hypothetical sample from the true conditional distribution:

$$Y_i^\# = l(X_i) + F^{-1}(U_i|X_i), \quad i = 1, \dots, n. \quad (2.6)$$

Note that the vectors (Y_1, \dots, Y_n) and $(Y_1^\#, \dots, Y_n^\#)$ are equally distributed given X_1, \dots, X_n . We are really interested in the exact values of $Y_i^\#$ and Y_i^* only when they are near the appropriate quantile, that is, only if $|U_i - p| < S_n \delta_n$. But then, by equation (2.1), Lemma 2.2.1 and the inverse function theorem, we have:

$$\begin{aligned} & \max_{i: |F^{-1}(U_i|X_i) - F^{-1}(p)| < S_n \delta_n} |F^{-1}(U_i|X_i) - \hat{F}^{-1}(U_i|X_i)| \\ &= \max_{i: |Y_i^\# - l(X_i)| < S_n \delta_n} |Y_i^\# - l(X_i) - Y_i^* + l_g(X_i)| = \mathcal{O}_p\{S_n \delta_n^{3/2}\}. \end{aligned} \quad (2.7)$$

Let now $q_{hi}(Y_1, \dots, Y_n)$ be the solution of the local quantile as given by (2.2) at X_i , with bandwidth h , i.e. $q_{hi}(Y_1, \dots, Y_n) \stackrel{\text{def}}{=} l_h(X_i)$ for data set $\{(X_i, Y_i)\}_{i=1}^n$. Note that by (2.3), if $|X_i - X_j| = \mathcal{O}(h)$, then

$$\max_{|X_i - X_j| < ch} |l_g(X_i) - l_g(X_j) - l(X_i) + l(X_j)| = \mathcal{O}_p(\delta_n) \quad (2.8)$$

Let l_h^* and $l_h^\#$ be the local bootstrap quantile and its coupled sample analogue. Then

$$\begin{aligned} l_h^*(X_i) - l_g(X_i) &= q_{hi}[\{Y_j^* - l_g(X_i)\}_{j=1}^n] \\ &= q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\}_{j=1}^n], \end{aligned} \quad (2.9)$$

while

$$l_h^\#(X_i) - l(X_i) = q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]. \quad (2.10)$$

From (2.7) – (2.10) we conclude that

$$\max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) + l(X_i)| = \mathcal{O}_p(\delta_n). \quad (2.11)$$

Based on (2.11), we obtain the following theorem (the proof is given in the appendix):

THEOREM 2.2.1 *If assumptions (A1 - A4) hold, then*

$$\sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) + l(x)| = \mathcal{O}_p(\delta_n) = \mathcal{O}_p(n^{-2/5}).$$

A number of replications of $l_h^*(x)$ can be used as the basis for simultaneous error bars because the distribution of $l_h^\#(x) - l(x)$ is approximated by the distribution of $l_h^*(x) - l_g(x)$, as Theorem 2.2.1 shows.

Although Theorem 2.2.1 is stated with a fixed bandwidth, in practice, to take care of the heteroscedasticity effect, we construct confidence bands with the width depending on the densities, which is motivated by the counterpart based on the asymptotic theory

as in Härdle and Song (2010). Thus we have the following corollary:

COROLLARY 2.2.1 *Let d_{α}^* be defined by $P^*(|l_h^*(x) - l_g(x)| > d_{\alpha}^*) = \alpha$, where P^* is the bootstrap distribution conditioned on the sample. If (A1)–(A4) hold, then the confidence interval $l_h(x) \pm d_{\alpha}^*$ has an asymptotic uniform coverage of $1 - \alpha$, in the sense that $P(\sup_{x \in J^*} |l_h(x) - l_g(x)| > d_{\alpha}^*) \rightarrow \alpha$.*

In practice we would use the approximate $(1 - \alpha) \times 100\%$ confidence band over \mathbb{R} given by

$$l_h(x) \pm \left[\hat{f}\{l_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_{\alpha}^*,$$

where d_{α}^* is based on the bootstrap sample (defined later) and $\hat{f}\{l_h(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$ with use of $f(y|x) = f(x, y)/f_X(x)$.

Below is the summary of the basic steps for the bootstrap procedure:

- 1) Given $(X_i, Y_i), i = 1, \dots, n$, compute the local quantile smoother $l_h(x)$ of Y_1, \dots, Y_n with bandwidth h and obtain residuals $\hat{\varepsilon}_i = Y_i - l_h(X_i)$, $i = 1, \dots, n$.
- 2) Compute the conditional edf:

$$\hat{F}(t|x) = \frac{\sum_{i=1}^n K_h(x - X_i) \mathbf{1}\{\hat{\varepsilon}_i \leq t\}}{\sum_{i=1}^n K_h(x - X_i)}$$

- 3) For each $i = 1, \dots, n$, generate random variables $\varepsilon_{i,b}^* \sim \hat{F}(t|x)$, $b = 1, \dots, B$ and construct the bootstrap sample $Y_{i,b}^*, i = 1, \dots, n, b = 1, \dots, B$ as follows:

$$Y_{i,b}^* = l_g(X_i) + \varepsilon_{i,b}^*.$$

- 4) For each bootstrap sample $\{(X_i, Y_{i,b}^*)\}_{i=1}^n$, compute l_h^* and the random variable

$$d_b \stackrel{\text{def}}{=} \sup_{x \in J^*} \left[\hat{f}\{l_h^*(x)|x\} \sqrt{\hat{f}_X(x)} |l_h^*(x) - l_g(x)| \right]. \quad (2.12)$$

where $\hat{f}\{l(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$.

- 5) Calculate the $(1 - \alpha)$ quantile d_{α}^* of d_1, \dots, d_B .
- 6) Construct the bootstrap uniform confidence band centered around $l_h(x)$, i.e. $l_h(x) \pm \left[\hat{f}\{l_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_{\alpha}^*$.

While bootstrap methods are well-known tools for assessing variability, more care must be taken to properly account for the type of bias encountered in nonparametric curve estimation. The choice of bandwidth is crucial here. In our experience the bootstrap works well with a rather crude choice of g , one may, however, specify g more precisely.

2.3 Bootstrap confidence bands in PLMs

Since the main role of the pilot bandwidth is to provide a correct adjustment for the bias, we use the goal of bias estimation as a criterion. Recall that the bias in the estimation of $l(x)$ by $l_h^\#(x)$ is given by

$$b_h(x) = \mathbb{E} l_h^\#(x) - l(x).$$

The bootstrap bias of the estimate constructed from the resampled data is

$$\hat{b}_{h,g}(x) = \mathbb{E} l_h^*(x) - l_g(x). \quad (2.13)$$

Note that in (2.13) the expected value is computed under the bootstrap estimation. The following theorem gives an asymptotic representation of the mean squared error for the problem of estimating $b_h(x)$ by $\hat{b}_{h,g}(x)$. It is then straightforward to find g to minimise this representation. Such a choice of g will make the quantiles of the original and coupled bootstrap distributions close to each other. In addition to the technical assumptions before, we also need:

5. l and f are four times continuously differentiable.
6. K is twice continuously differentiable.

THEOREM 2.2.2 *Under assumptions (A1 - A6), for any $x \in J^*$*

$$E \left[\left\{ \hat{b}_{h,g}(x) - b_h(x) \right\}^2 | X_1, \dots, X_n \right] \sim h^4 (C_1 g^4 + C_2 n^{-1} g^{-5}) \quad (2.14)$$

in the sense that the ratio between the RHS and the LHS tends in probability to 1 for some constants C_1, C_2 .

An immediate consequence of Theorem 2.2.2 is that the rate of convergence of g should be $n^{-1/9}$, see also Härdle and Marron (1991). This makes precise the previous intuition which indicated that g should slightly oversmooth. Under our assumptions, reasonable choices of h will be of the order $n^{-1/5}$ as in Yu and Jones (1998). Hence, (2.14) shows once again that g should tend to zero more slowly than h . Note that Theorem 2.2.2 is not stated uniformly over h . The reason is that we are only trying to give some indication of how the pilot bandwidth g should be selected.

2.3 Bootstrap confidence bands in PLMs

The case of multivariate regressors may be handled via a semiparametric specification of the quantile regression curve. More specifically we assume that with $x = (u, v)^\top \in \mathbb{R}^d$, $v \in \mathbb{R}$:

$$\tilde{l}(x) = u^\top \beta + l(v)$$

In this subsection we show how to proceed in this multivariate setting and how - based on Theorem 2.2.1 - a multivariate confidence band may be constructed. We first describe the numerical procedure for obtaining estimates of β and l , where l denotes - as in the earlier subsections - the one-dimensional conditional quantile curve. We then move on

2 Partial Linear Quantile Regression and Bootstrap Confidence Bands

to the theoretical properties. First note that the PLM quantile estimation problem can be seen as estimating (β, l) in

$$\begin{aligned} y &= u^\top \beta + l(v) + \varepsilon \\ &= \tilde{l}(x) + \varepsilon \end{aligned} \quad (2.15)$$

where the p -quantile of ε conditional on both u and v is 0.

In order to estimate β , let a_n denote an increasing sequence of positive integers and set $b_n = a_n^{-1}$. For each $n = 1, 2, \dots$, partition the unit interval $[0, 1]$ for v in a_n intervals $I_{ni}, i = 1, \dots, a_n$, of equal length b_n and let m_{ni} denote the midpoint of I_{ni} . In each of these small intervals $I_{ni}, i = 1, \dots, a_n$, $l(v)$ can be considered as being approximately constant, and hence (2.15) can be considered as a linear model. This observation motivates the following two stage estimation procedure:

- 1) A linear quantile regression inside each partition is used to estimate $\hat{\beta}_i, i = 1, \dots, a_n$. Their weighted mean yields $\hat{\beta}$. More exactly, consider the parametric quantile regression of y on $u, \mathbf{1}(v \in [0, b_n)), \mathbf{1}(v \in [b_n, 2b_n)), \dots, \mathbf{1}(v \in [1 - b_n, 1])$. That is, let

$$\psi(t) \stackrel{\text{def}}{=} (p-1)t\mathbf{1}(t < 0) + pt\mathbf{1}(t \geq 0).$$

Then let

$$\hat{\beta} = \arg \min_{\beta} \min_{l_1, \dots, l_{a_n}} \sum_{i=1}^n \psi\{Y_i - \beta^\top U_i - \sum_{j=1}^{a_n} l_j \mathbf{1}(V_i \in I_{ni})\}$$

- 2) Calculate the smooth quantile estimate as in (2.2) from $(V_i, Y_i - U_i^\top \hat{\beta})_{i=1}^n$, and name it as $\tilde{l}_h(v)$.

The following theorem states the asymptotic distribution of $\hat{\beta}$.

THEOREM 2.3.1 *If assumptions (A1) holds, for the above two stage estimation procedure, there exist positive definite matrices D, C , such that*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N\{0, p(1-p)D^{-1}CD^{-1}\} \text{ as } n \rightarrow \infty,$$

where $C = \text{plim}_{n \rightarrow \infty} C_n$ and $D = \text{plim}_{n \rightarrow \infty} D_n$ with $C_n = \frac{1}{n} \sum_{i=1}^n U_i^\top U_i$ and $D_n = \frac{1}{n} \sum_{j=1}^n f\{l(V_j)|V_i\} U_j^\top U_j$ respectively.

Note that $l(v), \tilde{l}_h(v)$ (quantile smoother based on $(v, y - u^\top \beta)$) and $\tilde{l}_h(v)$ can be treated as a zero (w.r.t. $\theta, \theta \in I$ where I is a possibly infinite, or possibly degenerate, interval

in \mathbb{R}) of the functions

$$\tilde{H}(\theta, v) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(v, \tilde{y}) \psi(\tilde{y} - \theta) d\tilde{y}, \quad (2.16)$$

$$\tilde{H}_n(\theta, v) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n K_h(v - V_i) \psi(\tilde{Y}_i - \theta), \quad (2.17)$$

$$\tilde{\tilde{H}}_n(\theta, v) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n K_h(v - V_i) \psi(\tilde{\tilde{Y}}_i - \theta), \quad (2.18)$$

where

$$\begin{aligned} \tilde{Y}_i &\stackrel{\text{def}}{=} Y_i - U_i^\top \beta \\ \tilde{\tilde{Y}}_i &\stackrel{\text{def}}{=} Y_i - U_i^\top \hat{\beta} = Y_i - U_i^\top \beta + U_i^\top (\beta - \hat{\beta}) \stackrel{\text{def}}{=} \tilde{Y}_i + Z_i. \end{aligned}$$

From Theorem 2.3.1 we know that $\hat{\beta} - \beta = \mathcal{O}_p(1/\sqrt{n})$ and $\|Z_i\|_\infty = \mathcal{O}_p(1/\sqrt{n})$. Under the following assumption, which are satisfied by exponential, and generalised hyperbolic distributions, also used in Härdle et al. (1988):

7. The conditional densities $f(\cdot|\tilde{y})$, $\tilde{y} \in \mathbb{R}$, are uniformly local Lipschitz continuous of order $\tilde{\alpha}$ (uL- $\tilde{\alpha}$) on J , uniformly in $\tilde{y} \in \mathbb{R}$, with $0 < \tilde{\alpha} \leq 1$, and $(nh)/\log n \rightarrow \infty$.

For some constant C_3 not depending on n , Lemma 2.1 in Härdle and Song (2010) shows a.s. as $n \rightarrow \infty$:

$$\sup_{\theta \in I} \sup_{v \in J^*} |\tilde{H}_n(\theta, v) - \tilde{H}(\theta, v)| \leq C_3 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\}.$$

Observing that $\sqrt{h/\log n} = o(1)$, we then have:

$$\begin{aligned} \sup_{\theta \in I} \sup_{v \in J^*} |\tilde{\tilde{H}}_n(\theta, v) - \tilde{H}(\theta, v)| &\leq \sup_{\theta \in I} \sup_{v \in J^*} |\tilde{H}_n(\theta, v) - \tilde{H}(\theta, v)| \\ &\quad + \underbrace{\sup_{\theta \in I} \sup_{v \in J^*} |\tilde{H}_n(\theta, v) - \tilde{\tilde{H}}_n(\theta, v)|}_{\leq \mathcal{O}_p(1/\sqrt{n}) \sup_{v \in J} |n^{-1} \sum K_h|} \\ &\leq C_4 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \end{aligned} \quad (2.19)$$

for a constant C_4 which can be different from C_3 . To show the uniform consistency of the quantile smoother, we shall reduce the problem of strong convergence of $\tilde{\tilde{l}}_h(v) - l(v)$, uniformly in v , to an application of the strong convergence of $\tilde{\tilde{H}}_n(\theta, v)$ to $\tilde{H}(\theta, v)$, uniformly in v and θ . For our result on $\tilde{\tilde{l}}_h(\cdot)$, we shall also require

8. $\inf_{v \in J^*} |\int \psi\{y - l(v) + \varepsilon\} dF(y|v)| \geq \tilde{q}|\varepsilon|$, for $|\varepsilon| \leq \delta_1$,

2 Partial Linear Quantile Regression and Bootstrap Confidence Bands

where δ_1 and \tilde{q} are some positive constants, see also Härdle and Luckhaus (1984). This assumption is satisfied if a constant \tilde{q} exists giving $f\{l(v)|v\} > \tilde{q}/p$, $x \in J$. Härdle and Song (2010) showed:

LEMMA 2.3.1 *Under assumptions (A7) and (A8), we have a.s. as $n \rightarrow \infty$*

$$\sup_{v \in J^*} |\tilde{l}_h(v) - l(v)| \leq C_5 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \quad (2.20)$$

with another constant C_5 not depending on n . If additionally $\tilde{\alpha} \geq \{\log(\sqrt{\log n}) - \log(\sqrt{nh})\}/\log h$, (2.20) can be further simplified to:

$$\sup_{v \in J^*} |\tilde{l}_h(v) - l(v)| \leq C_5 \{(nh/\log n)^{-1/2}\}.$$

Since the proof is essentially the same as Theorem 2.1 of the above mentioned reference, it is omitted here.

The convergence rate for the parametric part $\mathcal{O}_p(n^{-1/2})$ (Theorem 2.3.1) is smaller than the bootstrap approximation error for the nonparametric part $\mathcal{O}_p(n^{-2/5})$ as shown in Theorem 2.2.1. This makes the construction of uniform confidence bands for multivariate $x \in \mathbb{R}^d$ with a partial linear model possible.

PROPOSITION 2.3.1 *Under the assumptions (A1) - (A8), an approximate $(1 - \alpha) \times 100\%$ confidence band over $\mathbb{R}^{d-1} \times [0, 1]$ is*

$$u^\top \hat{\beta} + \tilde{l}_h(v) \pm \left[\hat{f}\{\tilde{l}_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_\alpha^*,$$

where $\hat{f}\{\tilde{l}_h(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$.

Note that here we actually only require the convergence rate of the parametric part, which is typically $\mathcal{O}_p(n^{-1/2})$, is smaller than the bootstrap approximation error for the nonparametric part $\mathcal{O}_p(n^{-2/5})$. This makes construction for the uniform confidence bands of more general semiparametric models possible instead of just the partial linear model shown here and similar results could be obtained easily.

2.4 A Monte Carlo study

This subsection is divided into two parts. First we concentrate on a univariate regressor variable x , check the validity of the bootstrap procedure together with settings in the specific example, and compare it with asymptotic uniform bands. Secondly we incorporate the partial linear model to handle the multivariate case of $x \in \mathbb{R}^d$.

Below is the summary of the simulation procedure:

- 1) Simulate $(X_i, Y_i), i = 1, \dots, n$ according to their joint pdf $f(x, y)$.

In order to compare with earlier results in the literature, we choose the joint pdf of bivariate data $\{(X_i, Y_i)\}_{i=1}^n$, $n = 1000$ as:

$$f(x, y) = f_{y|x}(y - \sin x)\mathbf{1}(x \in [0, 1]), \quad (2.21)$$

where $f_{y|x}(x)$ is the pdf of $N(0, x)$ with an increasing heteroscedastic structure. Thus the theoretical quantile is $l(x) = \sin(x) + \sqrt{x}\Phi^{-1}(p)$. Based on this normality property, all the assumptions can be seen to be satisfied.

- 2) Compute the local quantile smoother $l_h(x)$ of Y_1, \dots, Y_n with bandwidth h and obtain residuals $\hat{\varepsilon}_i = Y_i - l_h(X_i)$, $i = 1, \dots, n$.

If we choose $p = 0.9$, then $\Phi^{-1}(p) = 1.2816$, $l(x) = \sin(x) + 1.2816\sqrt{x}$. Set $h = 0.05$.

- 3) Compute the conditional edf:

$$\hat{F}(t|x) = \frac{\sum_{i=1}^n K_h(x - X_i)\mathbf{1}\{\hat{\varepsilon}_i \leq t\}}{\sum_{i=1}^n K_h(x - X_i)}$$

with the quartic kernel

$$K(u) = \frac{15}{16}(1 - u^2)^2, \quad (|u| \leq 1).$$

- 4) For each $i = 1, \dots, n$, generate random variables $\varepsilon_{i,b}^* \sim \hat{F}(t|x)$, $b = 1, \dots, B$ and construct the bootstrap sample $Y_{i,b}^*$, $i = 1, \dots, n$, $b = 1, \dots, B$ as follows:

$$Y_{i,b}^* = l_g(X_i) + \varepsilon_{i,b}^*,$$

with $g = 0.2$.

- 5) For each bootstrap sample $\{(X_i, Y_{i,b}^*)\}_{i=1}^n$, compute l_h^* and the random variable

$$d_b \stackrel{\text{def}}{=} \sup_{x \in J^*} \left[\hat{f}\{l_h^*(x)|x\} \sqrt{\hat{f}_X(x)} |l_h^*(x) - l_g(x)| \right]. \quad (2.22)$$

where $\hat{f}\{l(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$ with use of $f(y|x) = f(x, y)/f_X(x)$.

- 6) Calculate the $(1 - \alpha)$ quantile d_α^* of d_1, \dots, d_B .
- 7) Construct the bootstrap uniform confidence band centered around $l_h(x)$, i.e. $l_h(x) \pm \left[\hat{f}\{l_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_\alpha^*$.

Figure 2.1 shows the theoretical 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from the asymptotic theory and the confidence band from the bootstrap. The real 0.9 quantile curve is marked as the black dotted line.

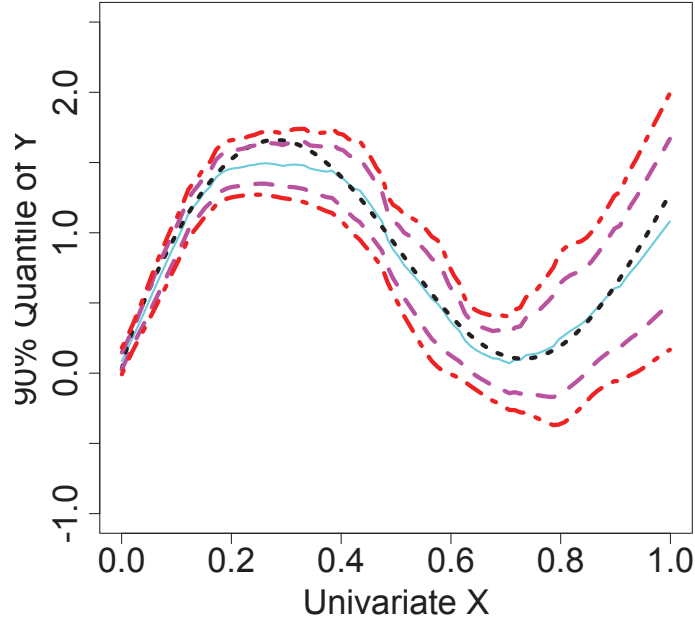


Figure 2.1: The real 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from asymptotic theory and confidence band from bootstrapping.

We then compute the classic local quantile estimate $l_h(x)$ (cyan solid) with its corresponding 95% uniform confidence band (magenta dashed) based on asymptotic theory according to Härdle and Song (2010). The 95% confidence band from the bootstrap is displayed as red dashed-dot lines. At first sight, the quantile smoother, together with two corresponding bands, all capture the heteroscedastic structure quite well, and the width of the bootstrap confidence band is similar to the one based on asymptotic theory in Härdle and Song (2010).

n	Cov. Prob.	Area
50	0.144 (0.642)	0.58 (1.01)
100	0.178 (0.742)	0.42 (0.58)
200	0.244 (0.862)	0.31 (0.36)

Table 2.1: Simulated coverage probabilities & areas of nominal asymptotic (bootstrap) 95% confidence bands with 500 repetition.

To compare the small sample performance and convergence rate of both methods, Table 2.1 presents the simulated coverage probabilities together with the calculated area of the 95% confidence band of the quantile smoother, for three sample sizes, $n = 50, 100$ and 200 . 500 simulation runs are carried out and for each simulation, 500 bootstrap samples are generated. From Table 2.1 we observe that, for the asymptotic

method, coverage probabilities improve with increasing sample size and the bootstrap method (shown inside brackets) obtains a significantly larger coverage probability than the asymptotic one, though still smaller than the nominal coverage, which results from the fact that quantile regression usually needs a larger sample size than mean regression and n here is quite moderate. It is also observed that the size of the bands decrease with increasing sample size. Overall, the bootstrap method displays a better convergence rate, while not sacrificing much on the width of the bands.

We now extend x to the multivariate case and use a different quantile function to verify our method. Choose $x = (u, v)^\top \in \mathbb{R}^d$, $v \in \mathbb{R}$, and generate the data $\{(U_i, V_i, Y_i)\}_{i=1}^n$, $n = 1000$ with:

$$y = 2u + v^2 + \varepsilon - 1.2816, \quad (2.23)$$

where u and v are uniformly distributed random variables in $[0, 2]$ and $[0, 1]$ respectively. ε has a standard normal distribution. The theoretical 0.9-quantile curve is $\tilde{l}(x) = 2u + v^2$. Since the choice of a_n is uncertain here, we test different choices of a_n for different n by simulation. To this end, we modify the theoretical model as follows:

$$y = 2u + v^2 + \varepsilon - \Phi^{-1}(p)$$

such that the real β is always equal to 2 no matter if p is 0.01 or 0.99. The result is displayed in Figure 2.2 for $n = 1000$, $n = 8000$, $n = 261148$ (number of observations for the data set used in the following application part). Different lines correspond to different a_n , i.e. $n^{1/3}/8$, $n^{1/3}/4$, $n^{1/3}/2$, $n^{1/3}$, $n^{1/3} \cdot 2$, $n^{1/3} \cdot 4$ and $n^{1/3} \cdot 8$. At first, it seems that the choice of a_n doesn't matter too much. To further investigate this, we calculate the SSE ($\sum_{i=1}^{99} \{\hat{\beta}(i/100) - \beta\}$) where $\hat{\beta}(i/100)$ denotes the estimate corresponding to the $i/100$ quantile. Results are displayed in Table 2.2. Obviously a_n has much less effect than n on SSE. Considering computational cost, which increases with a_n , and estimation performance, empirically we suggest $a_n = n^{1/3}$. Certainly this issue is far from settled and needs further investigations.

a_n	$n = 1000$	$n = 8000$	$n = 261148$
$n^{1/3}/8$			$3.6 * 10^{-3}$
$n^{1/3}/4$	$5.4 * 10^{-1}$	$4.0 * 10^{-2}$	$3.3 * 10^{-3}$
$n^{1/3}/2$	$6.1 * 10^{-1}$	$3.5 * 10^{-2}$	$3.2 * 10^{-3}$
$n^{1/3}$	$6.2 * 10^{-1}$	$3.6 * 10^{-2}$	$3.1 * 10^{-3}$
$n^{1/3} \cdot 2$	$8.0 * 10^{-1}$	$3.9 * 10^{-2}$	$2.9 * 10^{-3}$
$n^{1/3} \cdot 4$	$4.9 * 10^{-1}$	$3.6 * 10^{-2}$	$2.8 * 10^{-3}$
$n^{1/3} \cdot 8$			$3.4 * 10^{-3}$

Table 2.2: SSE of $\hat{\beta}$ with respect to a_n for different numbers of observations.

Thus for the specific model (2.23), we have $a_n = 10$, $\hat{\beta} = 1.997$, $h = 0.2$ and $g = 0.7$. In Figure 2.3 the theoretical 0.9 quantile curve with respect to v , and the 0.9 quantile

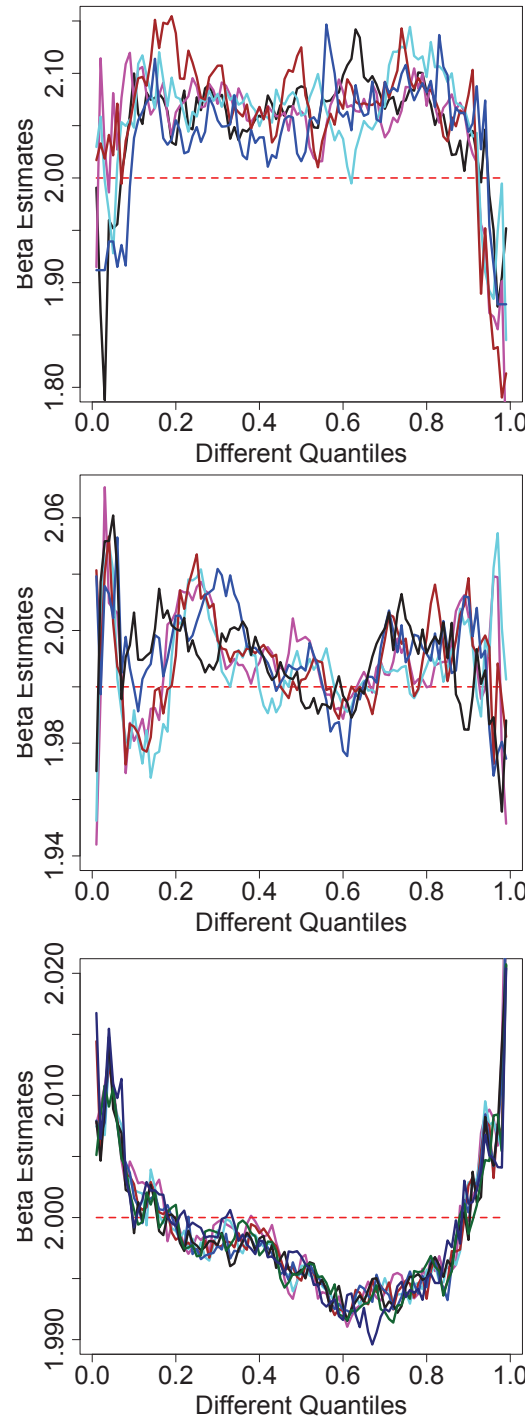


Figure 2.2: $\hat{\beta}$ with respect to different quantiles for different numbers of observations, i.e. $n = 1000$, $n = 8000$, $n = 261148$.

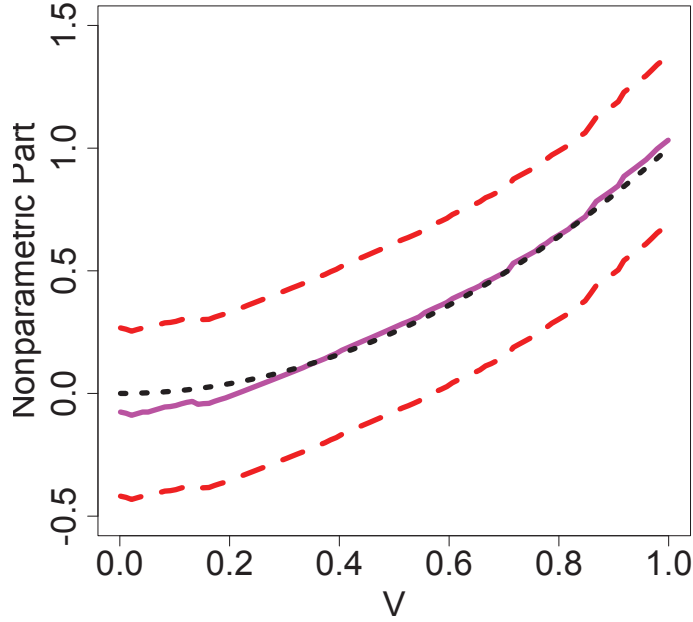


Figure 2.3: Nonparametric part smoothing, real 0.9 quantile curve with respect to v , 0.9 quantile smoother with corresponding 95% bootstrap uniform confidence band.

estimate with corresponding uniform confidence band are displayed. The real 0.9 quantile curve is marked as the black dotted line. We then compute the quantile smoother $l_h(x)$ (magenta solid). The 95% bootstrap uniform confidence band is displayed as red dashed lines and cover the true quantile curve quite well.

2.5 A labour market application

Our intuition of the effect of education on income is summarised by Day and Newburger (2002)'s basic claim: "At most ages, more education equates with higher earnings, and the payoff is most notable at the highest educational levels", which is actually from the point of view of mean regression. However, whether this difference is significant or not is still questionable, especially for different ends of the (conditionally) income distribution. To this end, a careful investigation of quantile regression is necessary. Since different education levels may reflect different productivity, which is unobservable and may also results from different ages, abilities etc, to study the labour market differential effect with respect to different education levels, a semiparametric partial linear quantile model is preferred, which can retain the flexibility of the nonparametric models for the age and other unobservable factors and ease the interpretation of the education factor.

We use the administrative data from the German National Pension Office (Deutsche Rentenversicherung Bund) for the following group: West Germany part, males aged 25 – 59, born between 1939 and 1942 who began receiving a pension in 2004 or 2005,

with at least 30 yearly uncensored observations, and thus in total, $n = 128429$ observations are available. We have the following three education categories: “low education”, “apprenticeship” and “university” for the variable u (assign them the numerical values 1, 2 and 3 respectively); the variable v is the age of the employee. “Low education” means without post-secondary education in Germany. “Apprenticeship” are part of Germany’s dual education system. Depending on the profession, they may work for three to four days a week in the company and then spend one or two days at a vocational school (Berufsschule). “University” in Germany also includes the technical colleges (applied universities). Since the level and structure of wages differs substantially between East and West Germany, we concentrate on West Germany only here (which we usually refer to simply as Germany). Our data have several advantages over the most often used German Socio-Economics Panel (GSOEP) data to analyze wages in Germany. Firstly, it is available for a much longer period, as opposed to from 1984 only for the GSOEP data. Secondly, more importantly, it has a much larger sample size. Thirdly, wages are likely to be measured much more precisely. Fourthly, we observe a complete earnings history from the individual’s first job until his retirement, therefore this is a true panel, not a pseudo-panel. There are also several drawbacks. For example, some very wealthy individuals are not registered in the German pension system, e.g. if the monthly income is more than some threshold (which may vary for different years due to the inflation effect), the individual has the right not to be included in the public pension system, and thus not recorded. Besides this, it is also right-censored at the highest level of earnings that are subject to social security contributions, so the censored observations in the data are only for those who actually decided to stay within the public system. Because of the combination of truncation and censoring, this thesis focuses on the uncensored data only, and we should not draw inferences from the very high quantile. Recently, similar data is also used to investigate the German wage structure as in Dustmann et al. (2009).

Following from Becker (1994)’s human capital mode, a log transformation is performed first on the hourly real wages (unit: EUR, in year 2000 prices). Figure 2.4 displays the boxplots for the “low education”, “apprenticeship” and “university” groups corresponding to different ages. In the data all ages (25 ~ 59) are reported as integers and are categorised in one-year groups. We rescaled them to the interval $[0, 1]$ by dividing by 40, with a corresponding bandwidth of 0.059 for the nonparametric quantile smoothers. This is equivalent to setting a bandwidth 2 in the original age data. This makes sense, because to detect whether a differential effect for different education levels exists, we compare the corresponding uniform confidence bands, i.e. differences indicate that the differential effect may exist for different education levels in the German labour market for that specific labour group.

Following an application of the partial linear model in Subsection 2.3, Figure 2.5 displays $\hat{\beta}$ with respect to different quantiles for 6, 13, 25 partitions, respectively. At first, the $\hat{\beta}$ curve is quite surprising, since it is not, as in mean regression, a positive constant, but rather varies a lot, e.g. $\hat{\beta}(0.20) = 0.026$, $\hat{\beta}(0.50) = 0.057$ and $\hat{\beta}(0.80) = 0.061$. Furthermore, it is robust to different numbers of partitions. It seems that the differences between the “low education” and “university” groups are different for different tails of the wage distribution. To judge whether these differences are significant, we use

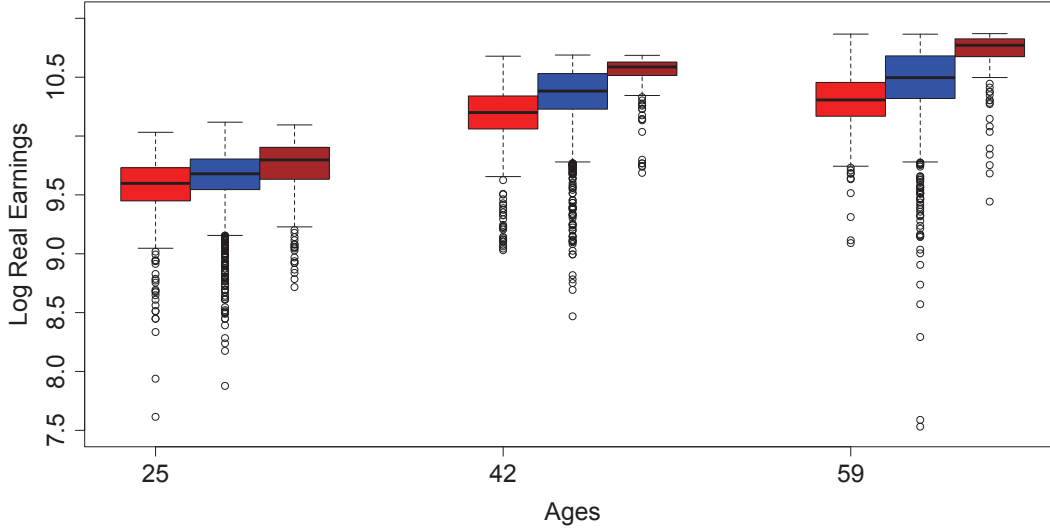


Figure 2.4: Boxplots for “low education”, “apprenticeship” & “university” groups corresponding to different ages.

the uniform confidence band techniques discussed in Subsection 2.2 which are displayed in Figure 2.6 - 2.8 corresponding to the 0.20, 0.50 and 0.80 quantiles respectively.

The 95% uniform confidence bands from bootstrapping for the “low education” group are marked as red dashed lines, while the ones for “apprenticeship” and “university” are displayed as blue dotted and brown dashed-dot lines, respectively. For the 0.20 quantile in Figure 2.6, the bands for “university”, “apprenticeship” and “low education” do not differ significantly from one another although they become progressively lower, which indicates that high education does not equate to higher earnings significantly for the lower tails of wages, while increasing age seems the main driving force. For the 0.50 quantile in Figure 2.7, the bands for “university” and “low education” differ significantly from one another although not from “apprenticeship”’s. However, for the 0.80-quantiles in Figure 2.8, all the bands differ significantly (except on the right boundary because of the non-parametric method’s boundary effect) resulting from the relatively large $\hat{\beta}(0.80) = 0.061$, which indicates that high education is significantly associated with higher earnings for the uppers tails of wages.

If we investigate the explanations for the differences in different tails of the income distribution, maybe the most prominent reason is the rapid development of technology, which has been extensively studied. The point is technology does not simply increase the demand for upper-end labour relative to that of lower-end labour, but instead asymmetrically affects the bottom and the top of the wage distribution, resulting in its strong asymmetry.

Conclusions from the point of view of quantile regression are consistent with the (grouped) mean regression’s, but in a careful way, i.e. we provide formal statistical tools to judge these uniformly. Partial linear quantile regression techniques, together with confidence bands, as developed in this thesis, display very interesting findings compared

with classic (mean) methods. Motivated by several key observations like the average income for female employees increase more than men's during the past few decades, partially because a better social welfare system means women can be more and more selective; and the "hollowing out" effect of employment, i.e. job growth in U.S., U.K. and continental Europe has increasingly been concentrated in the tails of the skill distribution over the last two decades, with disproportionate employment gains in high-wage, high-education occupations and low-wage, low-education occupations, further applications, for example to different genders, labour union status, nationalities and inequality analysis amongst other things will definitely bring more contributions to the differential analysis of the labour market.

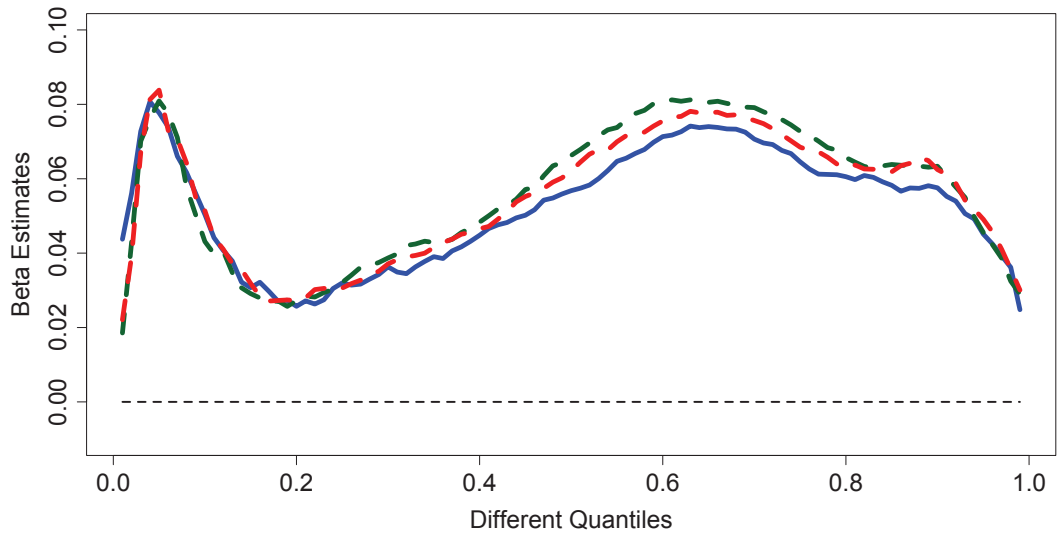


Figure 2.5: $\hat{\beta}$ corresponding to different quantiles with 6, 13, 25 partitions.

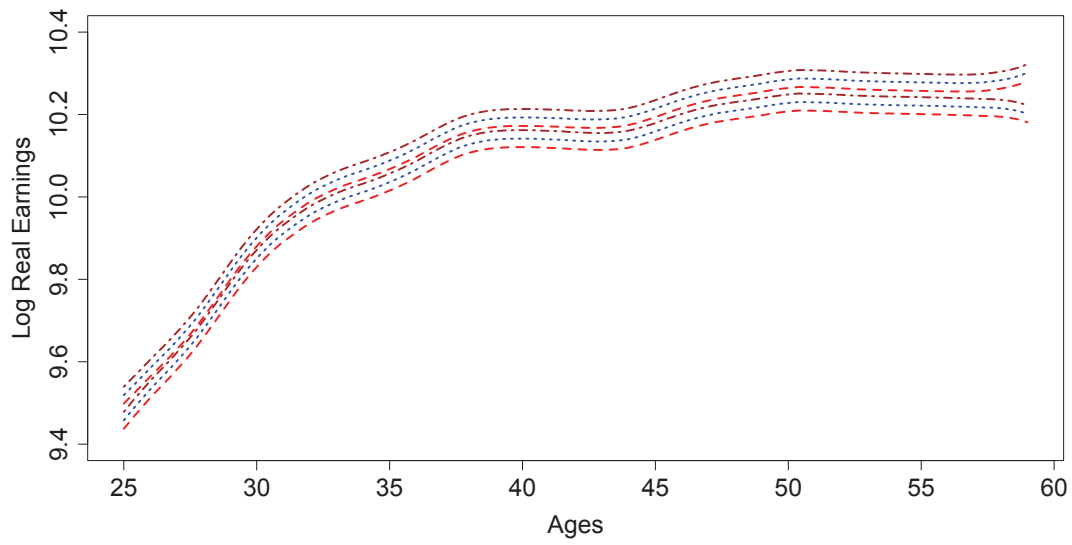


Figure 2.6: 95% uniform confidence bands for 0.05-quantile smoothers with 3 different education levels

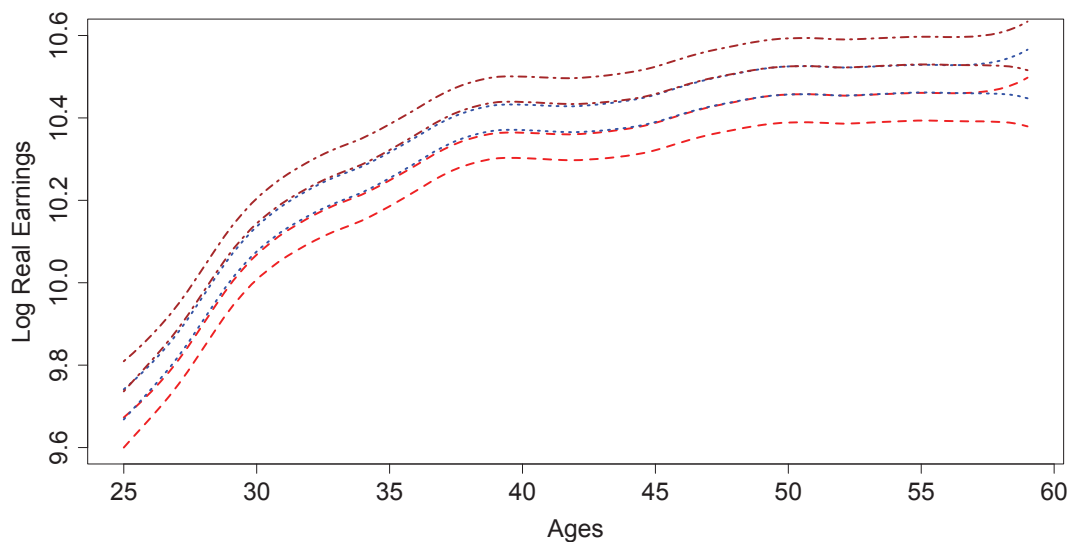


Figure 2.7: 95% uniform confidence bands for 0.50-quantile smoothers with 3 different education levels

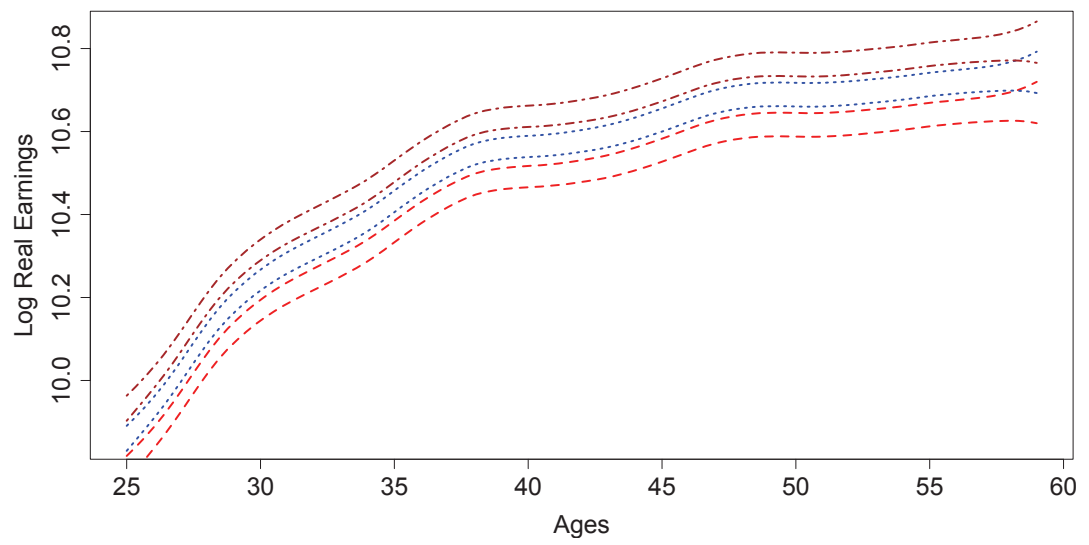


Figure 2.8: 95% uniform confidence bands for 0.99-quantile smoothers with 3 different education levels

2.6 Appendix

Proof of Theorem 2.2.1 We start by proving equation (2.7). Write first $\hat{F}^{-1}(U_i|X_i) = F^{-1}(U_i|X_i) + \Delta_i$. Fix any i such that $|F^{-1}(U_i|X_i)| \leq S_n \delta_n$, which, by equation (2.1), implies that $|U_i - p| < S_n \delta_n$. Lemma 2.2.1 gives:

$$\max_i |\hat{F}(S_n^2 \delta_n | X_i) - F(S_n^2 \delta_n | X_i)| = \mathcal{O}_p(\delta_n). \quad (2.24)$$

Together with $F(\pm S_n^2 \delta_n | X_i) = p \pm \mathcal{O}(S_n^2 \delta_n)$ again by equation (2.1), we have $\hat{F}(\pm S_n^2 \delta_n | X_i) = p \pm \mathcal{O}_p(S_n^2 \delta_n)$ and thus

$$\begin{aligned} \hat{F}(-S_n^2 \delta_n | X_i) &= p - \mathcal{O}_p(S_n^2 \delta_n) \\ &\leq p - S_n \delta_n < U_i < p + S_n \delta_n \\ &< p + \mathcal{O}_p(S_n^2 \delta_n) = \hat{F}(S_n^2 \delta_n | X_i). \end{aligned}$$

Since $\hat{F}(\cdot | X_i)$ is monotone non-decreasing, $|\hat{F}^{-1}(U_i | X_i)| \leq S_n^2 \delta_n$, which means, by $S_n^2 = S_n$,

$$|\hat{F}^{-1}(U_i | X_i)| \leq S_n \delta_n. \quad (2.25)$$

Apply now Lemma 2.2.1 again to equation (2.25), and obtain:

$$\begin{aligned} S_n \delta_n^{3/2} &\geq |\hat{F}\{\hat{F}^{-1}(U_i | X_i) | X_i\} - F\{\hat{F}^{-1}(U_i | X_i) | X_i\}| \\ &= |U_i - F\{F^{-1}(U_i | X_i) + \Delta_i | X_i\}| \\ &= |F\{F^{-1}(U_i | X_i) | X_i\} - F\{F^{-1}(U_i | X_i) + \Delta_i | X_i\}| \\ &\geq f_0(X_i) |\Delta_i| \end{aligned} \quad (2.26)$$

Hence $|\Delta_i| < S_n \delta_n^{3/2}$, and we summarise it as:

$$\max_{i: |F^{-1}(U_i | X_i) - F^{-1}(p)| < S_n \delta_n} |F^{-1}(U_i | X_i) - \hat{F}^{-1}(U_i | X_i)| = \mathcal{O}_p\{S_n \delta_n^{3/2}\}.$$

Beside the above approach, there is an alternative way. Note that $|\hat{F}^{-1}(U_i | X_i)| \leq |F^{-1}(U_i | X_i)| + |\Delta_i| \leq S_n \delta_n + |\Delta_i|$. Similar to inequality (2.26), by applying Lemma 2.2.1, we have $S_n \delta_n (|\Delta_i| + S_n \delta_n)^{1/2} \geq f_0(X_i) |\Delta_i|$. Solving this inequality w.r.t. $|\Delta_i|$ gives:

$$|\Delta_i| < \{S_n \delta_n^2 + (S_n \delta_n^2 + 4S_n \delta_n^3)^{1/2}\} / 2 = \mathcal{O}_p(S_n \delta_n^{3/2}),$$

which leads to the same conclusion.

To show equation (2.11), define

$$\begin{aligned} Z_{1j} &\stackrel{\text{def}}{=} Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i) \\ Z_{2j} &\stackrel{\text{def}}{=} Y_j^\# - l(X_j) + l(X_j) - l(X_i). \end{aligned}$$

2 Partial Linear Quantile Regression and Bootstrap Confidence Bands

Thus $q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\}_{j=1}^n]$ and $q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]$ can be seen as $l_h(X_i)$ for data sets $\{(X_i, Z_{1i})\}_{i=1}^n$ and $\{(X_i, Z_{2i})\}_{i=1}^n$ respectively. Similar to Härdle and Song (2010), they can be treated as a zero (w.r.t. θ , $\theta \in I$ where I is a possibly infinite, or possibly degenerate, interval in \mathbb{R}) of the functions

$$\tilde{G}_n(\theta, X_i) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n K_h(X_i - X_j) \psi(Z_{1j} - \theta), \quad (2.27)$$

$$\tilde{\tilde{G}}_n(\theta, X_i) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n K_h(X_i - X_j) \psi(Z_{2j} - \theta). \quad (2.28)$$

From (2.7) and (2.8), we have

$$\begin{aligned} \max_i \left| [\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\}_{j=1}^n] - [\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n] \right| \\ = \mathcal{O}_p\{S_n \delta_n^{3/2}\} + \mathcal{O}_p(\delta_n) = \mathcal{O}_p(\delta_n) \end{aligned} \quad (2.29)$$

Thus

$$\sup_{\theta \in I} \max_i |\tilde{G}_n(\theta, X_i) - \tilde{\tilde{G}}_n(\theta, X_i)| \leq \mathcal{O}_p(\delta_n) \max |n^{-1} \sum K_h| = \mathcal{O}_p(\delta_n)$$

To show the difference of the two quantile smoothers, we shall reduce the strong convergence of $q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\}_{j=1}^n] - q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]$, for any i , to an application of the strong convergence of $\tilde{G}(\theta, X_i)$ to $\tilde{\tilde{G}}_n(\theta, X_i)$, uniformly in θ , for any i . Under assumptions (A7) and (A8), in a similar spirit of Härdle and Song (2010), we get

$$\max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) - l(X_i)| = \mathcal{O}_p(\delta_n).$$

To show the supremum of the bootstrap approximation error, without loss of generality, based on assumption (A1), we reorder the original observations $\{X_i, Y_i\}_{i=1}^n$, such that $X_1 \leq X_2 \leq \dots \leq X_n$. First decompose:

$$\begin{aligned} \sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)| &= \max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) - l(X_i)| \\ &\quad + \max_i \sup_{x \in [X_i, X_{i+1}]} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)|. \end{aligned} \quad (2.30)$$

From assumption (A1) we know $l'(\cdot) \leq \lambda_1$ and $\max_i (X_{i+1} - X_i) = \mathcal{O}_p(S_n/n)$. By the mean value theorem, we conclude that the second term of (2.30) is of a lower order than the first term. Together with equation (2.11) we have

$$\begin{aligned} \sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)| \\ = \mathcal{O}\{\max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) - l(X_i)|\} = \mathcal{O}_p(\delta_n), \end{aligned}$$

which means that the supremum of the approximation error over all x is of the same order of the maximum over the discrete observed X_i . \square

Proof of Theorem 2.2.2. The proof of (2.14) uses methods related to those in the proof of Theorem 3 of Härdle and Marron (1991), so only the main steps are explicitly given. The first step is a bias-variance decomposition,

$$\mathbb{E} \left[\left\{ \hat{b}_{h,g}(x) - b_h(x) \right\}^2 \middle| X_1, \dots, X_n \right] = \mathcal{V}_n + \mathcal{B}_n^2 \quad (2.31)$$

where

$$\begin{aligned} \mathcal{V}_n &= \text{Var} \left[\hat{b}_{h,g}(x) \middle| X_1, \dots, X_n \right], \\ \mathcal{B}_n &= \mathbb{E} \left[\hat{b}_{h,g}(x) - b_h(x) \middle| X_1, \dots, X_n \right]. \end{aligned}$$

Following the uniform Bahadur representation techniques for quantile regression as in Theorem 3.2 of Kong et al. (2008), we have the following linear approximation for the quantile smoother as a local polynomial smoother corresponding to a specific loss function:

$$l_h^\#(x) - l(x) = L_n + \mathcal{O}_p(L_n),$$

where

$$L_n = \frac{n^{-1} \sum K_h(x - X_i) \psi \{Y_i - l(x)\}}{f \{l(x)|x\} f_X(x)}$$

for

$$\begin{aligned} \psi(u) &= p \mathbf{1}\{u \in (0, \infty)\} - (1 - p) \mathbf{1}\{u \in (-\infty, 0)\} \\ &= p - \mathbf{1}\{u \in (-\infty, 0)\}, \\ l(x - t) - l(x) &= l'(x)(-t) + l''(x)t^2 + \mathcal{O}(t^2), \\ \{l(x - t) - l(x)\}' &= l''(x)(-t) + l'''(x)t^2 + \mathcal{O}(t^2), \\ f(x - t) &= f(x) + f'(x)(-t) + f''(x)t^2 + \mathcal{O}(t^2), \\ f'(x - t) &= f'(x) + f''(x)(-t) + f'''(x)t^2 + \mathcal{O}(t^2), \\ \int K_h(t)tdt &= 0, \\ \int K_h(t)t^2dt &= h^2d_K, \\ \int K_h(t)\mathcal{O}(t^2)dt &= \mathcal{O}(h^2). \end{aligned}$$

Then we have

$$\mathcal{B}_n = \mathcal{B}_{n1} + \mathcal{O}(\mathcal{B}_{n1}),$$

where

$$\mathcal{B}_{n1} = \frac{\int K_g(x-t)\mathcal{U}_h(t)dt - \mathcal{U}_h(x)}{f_X(x)f\{l(x)|x\}}$$

for

$$\begin{aligned}\mathcal{U}_h(x) &= \int K_h(x-s)\psi\{l(s)-l(x)\}f(s)ds \\ &= \int K_h(t)\psi\{l(x-t)-l(x)\}f(x-t)dt.\end{aligned}$$

By differentiation, a Taylor expansion and properties of the kernel K (see assumption (A2)),

$$\begin{aligned}\mathcal{U}'_h(x) &= \int K_h(t)[\psi'\{l(x-t)-l(x)\}'f(x-t) \\ &\quad + \psi\{l(x-t)-l(x)\}f'(x-t)]dt.\end{aligned}$$

Here ψ' is the derivative of ψ except the 0 point, which actually does not matter since there is integration afterwards. Collecting terms, we get

$$\begin{aligned}\mathcal{U}'_h(x) &= \int K_h(t)\{\psi'l''(x)f'_X(x)t^2 + \psi'l'''f_X(x)t^2 \\ &\quad + af'''(x)t^2 + o(t^2)\}dt \\ &= \int K_h(t)\{C_0t^2 + o(t^2)\}dt = h^2d_K \cdot C_0 + o(h^2),\end{aligned}$$

where a is a constant with $|a| < 1$ and $C_0 = \psi'l''(x)f'_X(x) + \psi'l'''f_X(x) + af'''(x)$.

Hence, by another substitution and Taylor expansion, for the first term in the numerator of \mathcal{B}_{n1} , we have

$$\mathcal{B}_{n2} = g^2h^2(d_K)^2 \cdot C_0 + o(g^2h^2).$$

Thus, along almost all sample sequences,

$$\mathcal{B}_n^2 = C_1g^4h^4 + o(g^4h^4) \tag{2.32}$$

for $C_1 = (d_K)^4C_0^2/[f_X^2(x)f^2\{l(x)|x\}]$.

For the variance term, calculation in a similar spirit shows that

$$\mathcal{V}_n = \mathcal{V}_{n1} + o(\mathcal{V}_{n1}),$$

where

$$\mathcal{V}_{n1} = \frac{\int K_g^2(x-t)\mathcal{W}_h(t)dt - \{\int K_g(x-t)\mathcal{U}_h(t)dt\}^2f_X(x)f\{l(x)|x\}}{f_X(x)f\{l(x)|x\}}$$

for

$$\begin{aligned}\mathcal{W}_h(x) &= \int K_h^2(x-s) \psi \{l(s) - l(x)\}^2 f(s) ds \\ &= \int K_h^2(t) \psi \{l(x-t) - l(x)\}^2 f(x-t) dt.\end{aligned}$$

Hence, by Taylor expansion, collecting items and similar calculation, we have

$$\mathcal{V}_n = n^{-1} h^4 g^{-5} C_2 + o(n^{-1} h^4 g^{-5}) \quad (2.33)$$

for a constant C_2 . This, together with (2.31) and (2.32) completes the proof of Theorem 2.2.2. \square

Proof of Theorem 2.3.1. In case the function l is known, the estimate $\hat{\beta}_I$ is:

$$\hat{\beta}_I = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \psi \{Y_i - l(V_i) - U_i^\top \beta\}.$$

Since l is unknown, in each of these small intervals I_{ni} , $l(V_i)$ could be regarded as a constant $\alpha = l(m_{ni})$ for some i whose corresponding interval I_{ni} covers V_i . From assumption (A1), we know that $|l(V_i) - \alpha_i| \leq \lambda_1 b_n < \infty$. If we define our first step estimate $\hat{\beta}_i$ inside each small interval as

$$(\hat{\alpha}_i, \hat{\beta}_i) = \underset{\alpha, \beta}{\operatorname{argmin}} \sum \psi(Y_i - \alpha - U_i^\top \beta),$$

$|\{Y_i - l(V_i) - U_i^\top \beta\} - (Y_i - \alpha - U_i^\top \beta)| \leq \lambda_1 b_n < \infty$ indicates that we could treat $\hat{\beta}_i$ as $\hat{\beta}_I$ inside each partition. If we use d_i to denote the number of observations inside partition I_{ni} (based on the i.i.d. assumption as in assumption (A1), on average $d_i = n/a_n$). For each of the $\hat{\beta}_i$ inside interval I_{ni} , various parametric quantile regression literature, e.g. the convex function rule in Pollard (1991) and Knight (2001) yields

$$\sqrt{d_i}(\hat{\beta}_i - \beta) \xrightarrow{L} N\{0, p(1-p)D_i'^{-1}(p)C_i'D_i'^{-1}(p)\} \quad (2.34)$$

with the matrices $C_i' = d_i^{-1} \sum_{i=1}^{d_i} U_i^\top U_i$ and $D_i'(p) = d_i^{-1} \sum_{i=1}^{d_i} f\{l(V_i)|V_i\} U_i^\top U_i$.

To get $\hat{\beta}$, our second step is to take the weighted mean of $\hat{\beta}_1, \dots, \hat{\beta}_{a_n}$ as:

$$\begin{aligned}\hat{\beta} &= \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{a_n} d_i (\hat{\beta}_i - \beta)^2 \\ &= \sum_{i=1}^{a_n} d_i \hat{\beta}_i / n\end{aligned}$$

Please note that under this construction, $\hat{\beta}_1, \dots, \hat{\beta}_{a_n}$ are independent but not identical. Thus we intend to use the Lindeberg condition for the central limit theorem. To this

2 Partial Linear Quantile Regression and Bootstrap Confidence Bands

end, we use s_n^2 to denote $\text{Var}(\sum_{i=1}^{a_n} d_i \hat{\beta}_i / n)$, and we need to further check whether the following ‘‘Lindeberg condition’’ holds:

$$\lim_{a_n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^{a_n} \int_{(|d_i \hat{\beta}_i / n - \beta| > \varepsilon s_n)} (\hat{\beta}_i - \beta)^2 dF = 0, \quad \text{for all } \varepsilon > 0. \quad (2.35)$$

Since

$$\begin{aligned} \text{Var}(\sum_{i=1}^{a_n} d_i \hat{\beta}_i / n) &= \sum_i^{a_n} p(1-p) \left\{ \left[n/d_i \sum_{j=1}^{d_i} f\{l(V_j)|v\} U_j^\top U_j \right]^{-1} \right. \\ &\quad \times \sum_{i=1}^{d_i} U_i^\top U_i \left[n/d_i \sum_{j=1}^{d_i} f\{l(V_j)|v\} U_j^\top U_j \right]^{-1} \Big\} \\ &\approx p(1-p) \left[\sum_{j=1}^n f\{l(V_j)|v\} U_j^\top U_j \right]^{-1} \\ &\quad \times \sum_{i=1}^n U_i^\top U_i \left[\sum_{j=1}^n f\{l(V_j)|v\} U_j^\top U_j \right]^{-1} \\ &\stackrel{\text{def}}{=} \frac{1}{n} p(1-p) D_n^{-1} C_n D_n^{-1}, \end{aligned}$$

where $D_n = \frac{1}{n} \sum_{j=1}^n f\{l(V_j)|V_i\} U_j^\top U_j$ and $C_n = \frac{1}{n} \sum_{i=1}^n U_i^\top U_i$, together with the normality of $\hat{\beta}_i$ as in (2.34) and properties of the tail of the normal distribution, e.g. Exe. 14.3 – 14.4 of Borak et al. (2010), (2.35) follows.

Thus as $n, a_n \rightarrow \infty$ (although at a lower rate than n), together with $C = \text{plim}_{n \rightarrow \infty} C_n$, $D = \text{plim}_{n \rightarrow \infty} D_n$, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N\{0, p(1-p)D^{-1}CD^{-1}\}. \quad (2.36)$$

3 Generalized Dynamic Semiparametric Factor Models

3.1 Introduction

Modeling high-dimensional data is a challenging task in statistics especially when the data come in a dynamic context and are observed at different time points with changing structure and different sample sizes. Such modeling challenges appear in many different fields. In meteorology and agricultural economics, one of the primary interests is to study fluctuations of temperatures at different locations, for a recent summary, see Gleick et al. (2010). Such an analysis is essential for pricing weather derivatives and hedging weather risks, Odening et al. (2008). In neuro-economics, one uses (high dimensional) functional magnetic resonance imaging data (fMRI) to analyze the brain's response to certain (economics related) stimuli as well as identifying its activation area, Worsley et al. (2002). In financial engineering, one studies the dynamics of the implied volatility surface for risk management, calibration and pricing purposes, Fengler et al. (2007). Other examples and research fields for very large dimensional time series include empirical macroeconomics, Stock and Watson (2005); mortality analysis, Lee and Carter (1992); bond portfolio risk management or derivative pricing, Nelson and Siegel (1987) and Diebold and Li (2006); limit order book dynamics, Hall and Hautsch (2006); yield curves, Hautsch and Ou (2008). In the biostatistical field, we refer to Martinussen and Scheike (2000) for biomedical research; Kauermann (2000) for radiation treatment of prostate cancer; Gasser et al. (1983) for Electroencephalogram (EEG) analysis.

The modeling challenge for high dimensional time series is that there are both high dimensionality (in space) and dynamics (in time). One approach utilizes a factor type model, which allows low-dimensional representation of the data by separating high dimensionality and dynamics, see Forni et al. (2005), Giannone et al. (2005), Stock and Watson (2002a), Stock and Watson (2002b). In an orthogonal L -factor model, a J -dimensional random vector $Y_t = (Y_{t,1}, \dots, Y_{t,J})^\top$ can be represented as

$$Y_{t,j} = Z_{t,1}m_{1,j} + \dots + Z_{t,L}m_{L,j} + \varepsilon_{t,j}, \quad (3.1)$$

where $Z_{t,l}$ are common factors, $\varepsilon_{t,j}$ are errors and the coefficients $m_{l,j}$ are factor loadings. In the above described applications, the index $t = 1, \dots, T$ reflects the time evolution, and Y_t can be considered as a multidimensional not necessarily stationary time series. The study of the time behavior of the high-dimensional Y_t is then simplified to the modeling of $Z_t = (Z_{t,1}, \dots, Z_{t,L})^\top$, which is a more feasible task when $L \ll J$. In a variety of applications, one has explanatory variables $X_{t,j} \in \mathbb{R}^d$ at hand that may

3 Generalized Dynamic Semiparametric Factor Models

influence the factor loadings m_l . An important refinement of the model (3.1) is to incorporate the existence of observable covariates $X_{t,j}$. The factor loadings are then generalized to functions of $X_{t,j}$, so that the model (3.1) is generalized to:

$$\begin{aligned} Y_{t,j} &= \sum_{l=1}^L Z_{t,l} m_l(X_{t,j}) + \varepsilon_{t,j}, \quad 1 \leq j \leq J_t, \quad 1 \leq t \leq T. \\ &\stackrel{\text{def}}{=} Z_t^\top m(X_{t,j}) + \varepsilon_{t,j} \end{aligned} \quad (3.2)$$

where $Z_t = (Z_{t,1}, \dots, Z_{t,L})^\top$ (common factors) is an unobservable L -dimensional process (not necessarily stationary), m (factor loading functions) is an L -tuple (m_1, \dots, m_L) of unknown real-valued functions m_l defined on a subset of \mathbb{R}^d and $\varepsilon_{t,j}$ are errors. The variables $X_{1,1}, \dots, X_{T,J_T}$, $\varepsilon_{1,1}, \dots, \varepsilon_{T,J_T}$ are independent. Throughout the paper we assume that the $X_{t,j}$ are deterministic. The errors $\varepsilon_{t,j}$ are *i.i.d.*, have zero mean and finite second moments. Park et al. (2009) consider this model when Z_t is stationary and call it a dynamic semiparametric factor model (DSFM). For simplicity of notation, we assume that the covariates $X_{t,j}$ have support $[0, 1]^d$, and also that $J_t \equiv J$ do not depend on t unless otherwise specified.

The approximation (3.2) involves unknown “space functions” $m_l(\cdot)$ which in Park et al. (2009) are estimated via a B-Spline series:

$$m_l(x) = \sum_{k=1}^K a_{lk} \psi_k(x) \quad (3.3)$$

with a possibly multidimensional (as a tensor product of one dimensional) B-spline basis $\{\psi_k\}_{k=1}^K$. Using the $K \times J$ matrix $\Psi_t = \{\psi_1(x_t), \dots, \psi_K(x_t)\}^\top$ and the matrix $A = (a_{lk}), l = 1, \dots, L, k = 1, \dots, K$ we can rewrite (3.2) as $Y_t = Z_t^\top A \Psi_t + \varepsilon_t$. Expanding the time effect in a series leads us to modeling Z_t as a sum of basis functions as well:

$$Z_{tl} = \sum_{r=1}^R \gamma_{rl} u_r(t) \quad (3.4)$$

Putting (3.3) and (3.4) together we obtain (3.5) and (3.6), i.e. we observe $(X_{t,j}, Y_{t,j})$ for $j = 1, \dots, J_t$ and $t = 1, \dots, T$ such that

$$Y_{t,j} = \sum_{l=1}^L \sum_{r=1}^R u_r(t) \gamma_{rl} \sum_{k=1}^K a_{lk} \psi_k(X_{t,j}) + \varepsilon_{t,j} \quad (3.5)$$

$$Y_t^\top = \underbrace{U_t^\top \Gamma^*}_{Z_t^\top} \underbrace{A^* \Psi_t}_m + \varepsilon_t \stackrel{\text{def}}{=} U_t^\top \beta^{*\top} \Psi_t + \varepsilon_t. \quad (3.6)$$

Here $U_t^\top = (u_1(t), \dots, u_R(t))$ is a $1 \times R$ matrix with $u_r(t)$ as the pre-specified initial time basis, which we introduce to capture the global trend and periodic variations. $\Psi_t = (\psi_1(X_t), \dots, \psi_K(X_t))^\top$ is a $K \times J$ matrix with ψ_k a space basis function. Γ^*, A^*

and $\beta^{*\top}$ are $R \times L$, $L \times K$ and $R \times K$ (unknown) underlying coefficient matrices consisting of γ_{rl} , a_{lk} and β_{rk} respectively. For every β matrix, we introduce $\beta_r = (\beta_{rk}, 1 \leq k \leq K)$, that is, the column vector formed by the coefficients corresponding to the r -th time basis. Additionally we define $\|\beta\|_{2,1} = \sum_{r=1}^R \sqrt{\sum_{k=1}^K \beta_{rk}^2}$. Finally we set $\mathcal{R}(\beta) = \{r : \beta_r \neq 0\}$ and $M(\beta) = |\mathcal{R}(\beta)|$ where $|\mathcal{R}(\beta)|$ denotes the cardinality of set $\mathcal{R}(\beta)$. For sake of simplicity and convenience, we sometimes use $|\cdot|$ to denote the L_1 norm for vectors and $\|\cdot\|$ to denote the L_2 norm for vectors or the mixed $(2, 1)$ norm for matrices.

Since certainly not all initially included time basis are fully loading, to avoid over-parametrization in time, basis or variable selection is necessary, i.e. some β_r s will be shrunk to 0 equivalently. A popular variable selection method is Lasso, Tibshirani (1996). An extension for factor structured models is the group Lasso, Yuan and Lin (2006), in which the penalty term is a mixed $(2, 1)$ -norm of the coefficient matrix.

Under an additional Gaussian error assumption, we first show that this group Lasso type estimator enjoys sparsity inequalities (upper bounds on the prediction error and the distance between the estimator and the true regression matrix β^*) and variable selection properties. Finally, we show how our results can be extended to more general noise distributions, of which we only require the variance to be finite. Since the standard assumption on ε_t being independent is often not met in practice, we further extend our results into the dependent scenario. Since the original model (3.6) actually assumes that there is no randomness in time, we face some restrictions in practice. To this end, we consider an extension incorporating the stochasticity (in time) and call it a generalized dynamic semiparametric factor model (GDSFM). But it also poses an important question: is it justified, from an inferential point of view, to base further statistical inference on the detrended stochastic time series? We show that the difference of the inference based on the estimated time series and “true” unobserved time series is asymptotically negligible, which finally allows one to study the dynamics of the whole high-dimensional system with a low dimensional stochastic process representation together with the deterministic trend.

Another motivation of (3.4) (the expansion in time), is from the temperature analysis (across China over the past 50 years). Our data set is taken from Climatic Data Center (CDC), China Meteorological Administration (CMA), which contains daily observations from 159 weather stations across China (reduced from 202 after data cleaning) from Jan 1st, 1957 to Dec 31st, 2009, as can be seen from Figure 3.1 (left) (average over the 159 weather stations’ observations). Except the well known seasonality effect, we may expect a climate change related trend. If we take the moving average of 730 nearby days, which is $(159 \cdot 730)^{-1} \sum_{s=-354}^{+365} \sum_{j=1}^{159} Y_{t+s,j}$ with $Y_{t,j}$ being the temperature of the j th weather station at time t , Figure 3.1 (right) shows a “large period” (around 10 years between peaks) and an upward trend of the Chinese temperatures. $X_{t,j} = X_j$ is the three-dimensional geographical information of the j th weather station. Studying the dynamics of temperatures in various places simultaneously using a well calibrated GDSFM model will enable us to forecast temperatures in time and space.

Another motivation for this research is from neuro-economics. Understanding which part of our brain is activated during risky decisions and whether there is a significant

3 Generalized Dynamic Semiparametric Factor Models

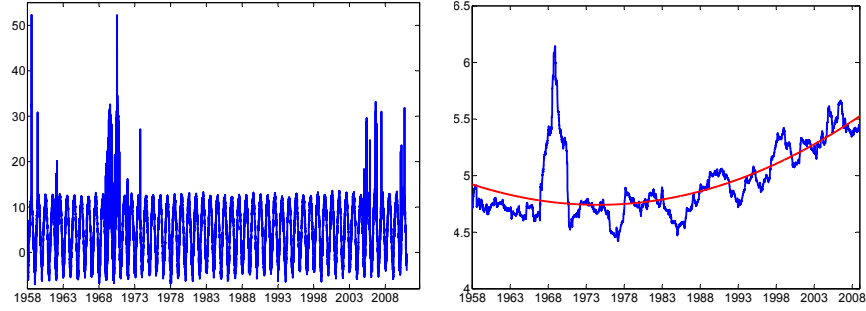


Figure 3.1: Temperatures of China from Jan 1st, 1957 to Dec 31st, 2009 (left) and the corresponding moving average (of 730 nearby days) view (right).

reaction to specific stimuli (neural processes underlying investment decisions) are important goals in neuroscience. We address this problem through the analysis of high dimensional, dynamic fMRI data recorded in an experiment (to be described in more detail later). The fMRI is a noninvasive technique of recording brain's signals on spatial area in a given time period (2.5 sec for our data set). One obtains a series of three-dimensional images of the blood-oxygen-level-dependent (BOLD) fMRI signals, when an exercised person is subject to certain stimuli related with financial decisions (periodically), where $Y_{t,j}$ is the BOLD value at voxel j and time t . $X_{t,j} = X_j$ is the three-dimensional geographical information of the j th voxel. An example of the images at one particular time point is presented in Figure 3.2.

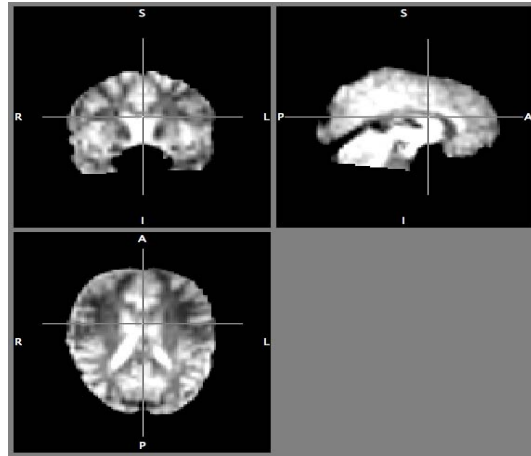


Figure 3.2: Typical fMRI data in one particular time point. The brightness corresponds to the strength of the observed signals.

The third motivation for this modeling approach (especially the space part) comes from financial engineering, i.e. the dynamics of the implied volatility surface (IVS) (although considered as stationary time series here), as is observed in Figure 3.3. The

IV is a volatility parameter that matches observed plain vanilla option prices with the ones given by the formula of Black and Scholes (1973), which is a key financial variable for trading, heading and the risk management of option portfolios. Figure 3.3 shows the “string” structure of the IV data obtained from European option prices on the German stock index DAX (ODAX) for two different days from the whole data set - intraday observations from Jan 1, 2004 to Dec 30, 2004 from Bloomberg. The volatility strings shift towards expiry, which is indicated by the bottom line in the figure. Moreover the shape of the IV strings is subject to stochastic deformation. Apart from the dynamic degeneration, one may also observe nonuniform frequency of the trades with significant greater market activities and the “smile” effect for the options closer to expiry or at-the-money. Fengler et al. (2007) first proposed to study the dynamics of the IV data, where $Y_{t,j}$ are the values of IV on the day t , and $X_{t,j}$ are the two-dimensional vectors of the moneyness and time-to-maturity, where the dimensionality J (number of transactions) depends also on t .

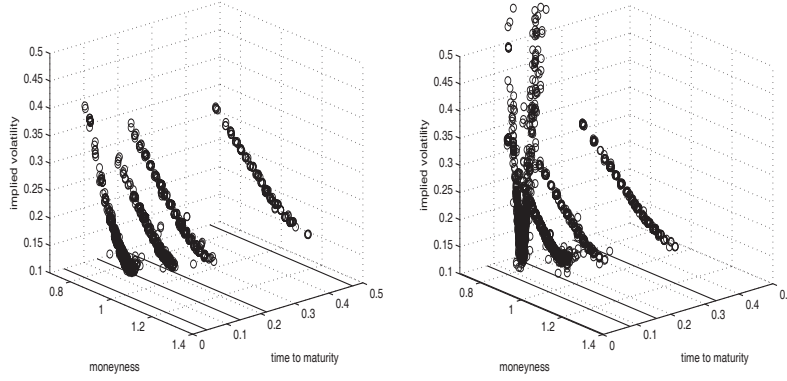


Figure 3.3: The typical IV data design on two different days. In the maturity direction observations appear in the discrete points for each particular day. Bottom solid lines indicate the observed maturities. Left panel: observations on 20040701, $J_t = 5606$. Right panel: observations on 20040819, $J_t = 8152$.

The rest of the article is organized as follows. In the next section we present the estimation of (3.6) to extract the complex deterministic trends of the nonstationary time series using the group Lasso type technique. Its properties under various situations are presented in Section 3.3. Section 3.4 considers the general framework incorporating the stochasticity (in time) together with the corresponding asymptotic analysis. In Section 3.5 we present the results of simulation studies that illustrate the theoretical findings. In Section 3.6 we apply the model to the temperature, IVS and fMRI data, where a panel version of (3.6) is also presented. All technical proofs are sketched in Section 3.7.

3.2 Methodology

3.2.1 Choice of Time Basis

To capture the global trend in time, one may use an orthogonal Legendre polynomial basis: $u_1(t) = 1/C_1, u_2(t) = t/C_2, u_3(t) = (3t^2 - 1)/C_3, \dots$ (throughout this paper, C_i are generic constants). The rescaling is made here such that $\sum_{t=1}^T u_r^2(t)/C_r^2 = 1$. To capture periodic variations, we could use Fourier series, $u_4(t) = \sin(2\pi t/p)/C_4, u_5(t) = \cos(2\pi t/p)/C_5, u_6(t) = \sin\{2\pi t/(p/2)\}/C_6, u_7(t) = \cos\{2\pi t/(p/2)\}/C_7, \dots$ with the given the period p . For example, in the fMRI application, we know that $p = 11.8$ (29.5s per trial & 2.5s per scan) and in the weather application, $p_1 = 365, p_2 = 365 \cdot 10$.

3.2.2 Choice of Space Basis

There are various choices for a space basis. For example, Park et al. (2009) use a series estimator as described in (3.3). However, it has some disadvantages. Firstly, since the B-spline basis $\{\psi_k\}_{k=1}^K$ is possibly multidimensional ($d > 1$), it is constructed as a tensor product of one dimensional ones. When $d \geq 3$, this may lead to quite large K , e.g. $K = 9 \times 9 \times 5 = 405$ in the fMRI application. More importantly, since the knots of the B-spline are equal-spaced, it could not capture some special structure, e.g. the “smile” effect in the IVS modeling when the options are close to the maturity, as can be seen in Figure 3.4 from Park et al. (2009) (adaptive choice of the knots of the B-splines may solve this problem, but it is omitted here since not primary interest).

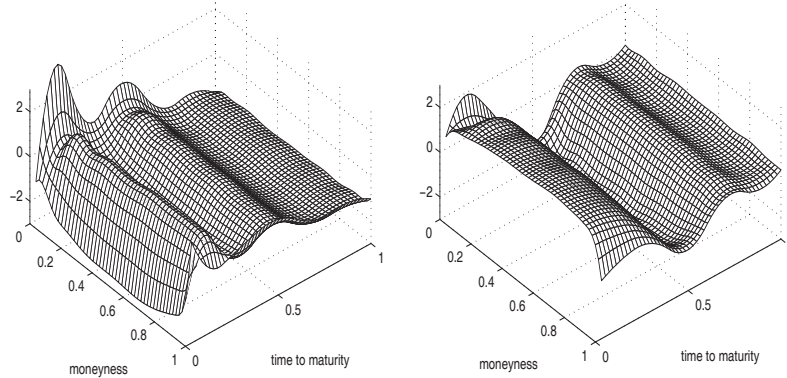


Figure 3.4: Space basis using the series estimator for the IVS modeling.

To this end, we propose a data driven method to estimate the space basis $\psi_1(x), \dots, \psi_K(x)$, motivated by Hall et al. (2006), which combines smoothing techniques with ideas related to functional principal component analysis. We summarize the basic steps as follows:

- 1 Estimate the covariance operator. Write $X_{tj} = (X_{tj}^1, \dots, X_{tj}^d)$, $u = (u^1, \dots, u^d)$ and $v = (v^1, \dots, v^d)$ (same for $b, \hat{b}, b_1, \hat{b}_1, b_2$ and \hat{b}_2). Given $u \in [0, 1]^d$, let h_μ and

h_ϕ denote bandwidths, which could be selected as in the usual local polynomial regression setup and select $(\hat{a}, \hat{b}) = (a, b)$ to minimize

$$\sum_{t=1}^T \sum_{j=1}^{J_t} \{Y_{tj} - a - \sum_{c=1}^d b^c(u^c - X_{tj}^c)\}^2 K\left(\frac{X_{tj} - u}{h_\mu}\right),$$

and take $\hat{\mu}(u) = \hat{a}$. Then, given $u, v \in [0, 1]^d$, choose $(\hat{a}_0, \hat{b}_1, \hat{b}_2) = (a_0, b_1, b_2)$ to minimize

$$\begin{aligned} & \sum_{t=1}^T \sum_{1 \leq j \neq k \leq J_t} \{Y_{tj}Y_{tk} - a_0 - \sum_{c=1}^d b_1^c(u^c - X_{tj}^c) - \sum_{c=1}^d b_2^c(v^c - X_{tk}^c)\}^2 \\ & \times K\left(\frac{X_{tj} - u}{h_\phi}\right) K\left(\frac{X_{tk} - v}{h_\phi}\right). \end{aligned}$$

Denote \hat{a}_0 by $\hat{\phi}(u, v)$ and construct $\hat{\mu}(v)$ similarly with $\hat{\mu}(u)$. The estimate of the covariance operator is thus:

$$\hat{\psi}(u, v) = \hat{\phi}(u, v) - \hat{\mu}(u)\hat{\mu}(v).$$

Since the covariance operator is $J \times J$, where J could be very large, to get its consistent estimates, various large covariance matrices regularization techniques, e.g. banding, Bickel and Levina (2008a) and thresholding, Bickel and Levina (2008b), could be further used.

- 2 Compute the principal space basis. Given the estimated operator, compute the largest K eigenvalues and corresponding orthonormal eigenfunctions as the basis $\psi_1(X_{t,j}), \dots, \psi_K(X_{t,j})$ ($\Psi_t \Psi_t^\top / J_t = I_K$ is thus valid). Computational methods could be found, for example, in Section 8.4 of Ramway and Silverman (2005), where practical features regarding the operator-eigenfunction implementation are discussed in detail.

3.2.3 Estimation Procedure

We have now accumulated sufficient information to introduce the estimation method, which is summarized as below:

- 1 Find significantly loaded time basis functions by the group Lasso technique by minimizing:

$$\min_{\beta} (JT)^{-1} \sum_{t=1}^T \left(Y_t^\top - U_t^\top \beta^\top \Psi_t \right) \left(Y_t^\top - U_t^\top \beta^\top \Psi_t \right)^\top + 2\lambda \|\beta\|_{2,1}. \quad (3.7)$$

- 2 Split the joint matrix $\hat{\beta}\hat{\beta}^\top$ into 2 separate coefficient matrices $\hat{\Gamma}, \hat{A}$ by taking $\hat{\Gamma}$ as the L eigenvectors of $\hat{\beta}\hat{\beta}^\top$ with respect to the L largest eigenvalues, and $\hat{A} = \hat{\Gamma}^\top \hat{\beta}$.

3 Generalized Dynamic Semiparametric Factor Models

To select K and L here, we could use either the classic “90%” rule in principal component analysis or the “explained variance” type selection method. Alternatively we could also sequentially test the size of the eigenvalues. But since it goes beyond the scope of this paper, we will therefore study its theoretical properties in a separate paper.

In order to study the statistical properties of this estimator, it is useful to derive some optimality condition for a solution of (3.7). Our implementation of the group Lasso-type estimator comes from Yuan and Lin (2006), which is an extension of the shooting algorithm of Fu (1998) for the lasso. As a direct consequence of the Karush-Kuhn-Tucker conditions, we have a necessary and sufficient condition for $\hat{\beta}$ to be a solution to expression (3.7) is

$$(JT)^{-1} \sum_{t=1}^T \{\Psi_t(Y_t - \Psi_t^\top \hat{\beta} U_t) U_t^\top\}_r = \lambda \frac{\hat{\beta}_r}{\|\hat{\beta}_r\|}, \text{ if } \hat{\beta}_r \neq 0 \quad (3.8)$$

$$(JT)^{-1} \left\| \sum_{t=1}^T \{\Psi_t(Y_t - \Psi_t^\top \hat{\beta} U_t) U_t^\top\}_r \right\| \leq \lambda, \quad \text{if } \hat{\beta}_r = 0 \quad (3.9)$$

Recall that $\Psi_t \Psi_t^\top / J = I_K$. It can be easily verified that the solution to (3.8) and (3.9) is

$$\hat{\beta}_r = \left(1 - \lambda / \|S_r\|\right)_+ S_r, \quad (3.10)$$

where $S_r = \sum_{t=1}^T \{\Psi_t(Y_t - \Psi_t^\top \hat{\beta}_{-r} U_t) U_t^\top\}_r$, with $\hat{\beta}_{-r} = (\hat{\beta}_1, \dots, \hat{\beta}_{r-1}, 0, \hat{\beta}_{r+1}, \dots, \hat{\beta}_R)$. The solution to expression (3.7) can therefore be obtained by iteratively applying equation (3.10) to $r = 1, \dots, R$. We choose the ordinary least square estimate $\hat{\beta}_{OLS}$ as the initial value, with which usually a reasonable convergence tolerance is reached within 5 iterations. However, the computational burden increases dramatically as the number of initial basis increases.

Since the group Lasso type estimates depend on the unknown tuning parameter parameter λ , which needs to be estimated, to select the final models on the solution paths of the group selection methods, we introduce an easily computable C_p -type criterion as in Yuan and Lin (2006). The solution path is computed by evaluating on 100 equally spaced λ 's between 0 and $\lambda_{max} = \max_r \|\sum_t \Psi_t Y_t U_{tr}\| / \sqrt{K}$. We select the λ minimizing

$$\begin{aligned} C_p(\lambda) &= \frac{\sum_t \|Y_t^\top - U_t^\top \hat{\beta}^\top \Psi_t\|^2}{\tilde{\sigma}^2} - JT + 2df \\ \tilde{\sigma}^2 &= \frac{\sum_t \|Y_t^\top - U_t^\top \hat{\beta}_{OLS}^\top \Psi_t\|^2}{JT - df} \\ df &= \sum_r \mathbf{1}\{\|\hat{\beta}_r\| > 0\} + \sum_r \frac{\|\hat{\beta}_r\|}{\|\hat{\beta}_{OLS}\|} (K - 1) \end{aligned}$$

Empirical evidence suggests that this approximation works fairly well. In our experience, the performance of this approximate C_p -criterion is generally comparable with

that of computationally much more expensive (especially for the high-dimensional data) fivefold cross-validation, as already noted in Yuan and Lin (2006).

3.3 Estimates' Properties

In this section, we first study the properties of this estimator as defined in (3.6) when the errors ε_t are Gaussian. Our main results concern upper bounds on the prediction error and the distance between the estimator and the true matrix β^* , (Theorem 3.3.1). The techniques of proofs are closely build upon those of Lounici et al. (2009), Bickel et al. (2008) and Lounici (2008). In Theorem 3.3.2 we discuss how our results can be extended to more general noise distribution, of which we only require the variance to be finite. Since the standard assumption on ε_t being independent is often not met in practice, in Theorem 3.3.3, we further extend our results into the dependent scenario.

LEMMA 3.3.1 *Consider the model (3.6) for $R \geq 2$ and $T, J \geq 1$. Assume that the random vectors $\varepsilon_1, \dots, \varepsilon_T$ are i.i.d. Gaussian with zero mean and covariance matrix $\sigma^2 I_{J \times J}$, $\Psi_t \Psi_t^\top / J = I_K$, $\sum_{t=1}^T U_t^\top U_t / R = 1$, and $M(\beta^*) \leq s$. Let*

$$\lambda = \frac{2\sigma}{\sqrt{JT}} \left(1 + A \log R / \sqrt{T}\right)^{1/2},$$

where $A > 8$ and let $q = \min(A \log R, \sqrt{T})$. Then with probability at least $1 - R^{1-q}$, for any solution $\hat{\beta}$ of problem (3.7) and $\forall \beta$ we have:

$$\begin{aligned} & (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\hat{\beta} - \beta^*) U_t\|^2 + \lambda \|\hat{\beta} - \beta\|_{2,1} \\ & \leq (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\beta - \beta^*) U_t\|^2 + 4\lambda \sum_{r \in \mathcal{R}(\beta)} \|\hat{\beta}_r - \beta_r\|, \end{aligned} \quad (3.11)$$

$$(JT)^{-1} \max_{1 \leq r \leq R} \left\| \sum_{t=1}^T \{\Psi_t \Psi_t^\top (\hat{\beta} - \beta^*) U_t U_t^\top\}_r \right\| \leq \frac{3}{2} \lambda, \quad (3.12)$$

and

$$M(\hat{\beta}) \leq \frac{4\phi_{\max}^2}{\lambda^2 T^2} \|\hat{\beta} - \beta^*\|_2^2, \quad (3.13)$$

where ϕ_{\max} is the maximum eigenvalue of the matrix $\sum_{t=1}^T U_t U_t^\top$.

Before stating the first main result of this section, we make the following assumption first.

ASSUMPTION 3.3.1 *There exists a positive number $\kappa = \kappa(s)$ such that*

$$\min \left\{ \frac{\sum_t \|\Psi_t^\top \Delta U_t\|}{\sqrt{J} \|\Delta_{\mathcal{R}}\|} : |\mathcal{R}| \leq s, \Delta \in \mathbb{R}^{K \times R} \setminus \{0\}, \right. \\ \left. \|\Delta_{\mathcal{R}^c}\|_{2,1} \leq 3 \|\Delta_{\mathcal{R}}\|_{2,1} \right\} \geq \kappa,$$

where \mathcal{R}^c denotes the complement of the set of indices \mathcal{R} , $\Delta_{\mathcal{R}}$ denotes the matrix formed by stacking the rows of matrix Δ w.r.t. row index set \mathcal{R} .

Assumption 3.3.1 is essentially a restriction on the eigenvalues of U_t as a function of sparsity s . It actually requires the initially involved time basis not to be too dependent, which is naturally satisfied by the orthogonal polynomials and Fourier series. Low sparsity means that s is big and therefore κ is small. $\kappa(s)$ is thus a decreasing function of s . For this reason we sometimes refer to it as Assumption RE(s), see also Bickel et al. (2008), but note that in their paper l_1 norms are used.

THEOREM 3.3.1 *Assume all conditions in Lemma 3.3.1 still hold and add Assumption 3.3.1. Then with probability at least $1 - R^{1-q}$, for any solution $\hat{\beta}$ of (3.7):*

$$(JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\hat{\beta} - \beta^*) U_t\|^2 \leq 64\sigma^2 s (1 + A \log R / \sqrt{T}) / (\kappa^2 J), \quad (3.14)$$

$$T^{-1/2} \|\hat{\beta} - \beta^*\|_{2,1} \leq 32\sigma s \sqrt{1 + A \log R / \sqrt{T}} / (\kappa^2 \sqrt{J}), \quad (3.15)$$

and

$$M(\hat{\beta}) \leq 64\phi_{\max}^2 s / \kappa^2 \quad (3.16)$$

Note that Theorem 3.3.1 is valid for any fixed J, R, T and therefore yields non-asymptotic bounds. We could see that dependence on the number of initially specified time basis R can be made negligible for large T . Additionally when the true coefficient matrix β^* 's sparsity level is low (s large, κ small, s/κ^2 large), all the three bounds get larger and the number of nonzero rows of estimated one $\hat{\beta}^\top$ is larger too correspondingly.

From now on, we only assume that the random variables ε_{tj} are independent with zero mean and finite variance $E(\varepsilon_{tj}^2) \leq \sigma^2$. In this case the results remain similar to those of the previous theorem, though the concentration effect is weaker. We use the following mild technical assumption.

ASSUMPTION 3.3.2 *The matrices Ψ_t and U_t are such that*

$$(JT)^{-1} \sum_{t=1}^T \sum_{j=1}^J \left(\max_r \left| \sum_{k=1}^K \Psi_{tkj} U_{tr} \right| \right)^2 \leq C,$$

for a constant $C > 0$.

THEOREM 3.3.2 Consider the DSFM (3.6) for $R \geq 3$ and $T, J \geq 1$. Assume that the random vectors $\varepsilon_1, \dots, \varepsilon_T$ are independent with zero mean and finite variance $E(\varepsilon_{tj}^2) \leq \sigma^2$, $\Psi_t \Psi_t^\top / J = I_K$, $\sum_{t=1}^T U_t^\top U_t / R = 1$, and $M(\beta^*) \leq s$. Let also Assumption 3.3.2 be satisfied. Furthermore let κ be defined as in Assumption 3.3.1 and ϕ_{\max} is the maximum eigenvalue of the matrix $\sum_{t=1}^T U_t U_t^\top$. Let

$$\lambda = \sigma \sqrt{(\log R)^{1+\delta} / (JT)}, \quad \delta > 0.$$

Then with probability at least $1 - (2e \log R - e)C / (\log R)^{1+\delta}$, for any solution $\hat{\beta}$ of (3.7) we have:

$$(JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\hat{\beta} - \beta^*) U_t\|^2 \leq 16\sigma^2 s (\log R)^{1+\delta} / (\kappa^2 J)$$

$$T^{-1/2} \|\hat{\beta} - \beta^*\|_{2,1} \leq 16\sigma s \sqrt{(\log R)^{1+\delta} / (\kappa^2 \sqrt{J})}$$

and

$$M(\hat{\beta}) \leq 64\phi_{\max}^2 s / \kappa^2$$

Since the standard assumption on ε_t being independent is often not met in practice, it is important to understand how the proposed estimator behaves under dependent error terms. As far as we know, our result is the first attempt with dependent error terms for (group) Lasso variable selection techniques. The other effort of getting rid of the independence assumption could be found in Jia et al. (2009), where they consider a sparse Poisson-like model. Before moving on, similar to Janson (2004), we introduce the following definitions first.

Given a set \mathcal{T} and random variables V_t , $t \in \mathcal{T}$, we say:

- A subset \mathcal{T}' of \mathcal{T} is *independent* if the corresponding random variables $\{V_t\}_{t \in \mathcal{T}'}$ are independent.
- A family $\{\mathcal{T}_j\}_j$ of subsets of \mathcal{T} is a *cover* of \mathcal{T} if $\bigcup_j \mathcal{T}_j = \mathcal{T}$.
- A family $\{(\mathcal{T}_j, w_j)\}_j$ of pairs (\mathcal{T}_j, w_j) , where $\mathcal{T}_j \subseteq \mathcal{T}$ and $w_j \in [0, 1]$ is a *fractional cover* of \mathcal{T} if $\sum_j w_j \mathbf{1}_{\mathcal{T}_j} \geq \mathbf{1}_{\mathcal{T}}$, i.e. $\sum_{j: t \in \mathcal{T}_j} w_j \geq 1$ for each $t \in \mathcal{T}$.
- A (fractional) cover is *proper* if each set \mathcal{T}_j in it is independent.
- $\mathcal{X}(\mathcal{T})$ is the size of the smallest proper cover of \mathcal{T} , i.e. the smallest m such that \mathcal{T} is the union of m independent subsets.
- $\mathcal{X}^*(\mathcal{T})$ is the minimum of $\sum_j w_j$ over all proper fractional covers $\{(\mathcal{T}_j, w_j)\}_j$.

Note that, in spite of our notation, $\mathcal{X}(\mathcal{T})$ and $\mathcal{X}^*(\mathcal{T})$ depend not only on \mathcal{T} but also on the family $\{V_t\}_{t \in \mathcal{T}}$. Note further that $\mathcal{X}^*(\mathcal{T}) \geq 1$ (unless $\mathcal{T} = \emptyset$) and that $\mathcal{X}^*(\mathcal{T}) = 1$

3 Generalized Dynamic Semiparametric Factor Models

if and only if the variables $V_t, t \in \mathcal{T}$ are independent, i.e. $\mathcal{X}^*(\mathcal{T})$ is a measure of the dependence structure of $\{V_t\}_{t \in \mathcal{T}}$. For example, if V_t just depends on V_{t-1} but independent of all $V_s, s < t-1$, e.g. $\text{AR}(1)$, $\mathcal{X}^*(\mathcal{T}) = 2$.

We use the following mild technical assumption similar to Assumption 3.3.2.

ASSUMPTION 3.3.3 *The matrices Ψ_t and U_t and random variables ε_t are such that*

$$(J^{-1} \sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr})^2 \leq b_t^2 \quad \text{with a high probability}$$

$$E(JT)^{-1} \left\{ \sum_{t=1}^T \left(\sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\}^{1/2} \leq \frac{C'}{\sqrt{T}}.$$

for $\forall r$ and some constants $b_t, C' > 0, t = 1, \dots, T$. Note that dropping the sub-index r for all constants here does not matter, since they could be taken as the maximum of all corresponding constants over different r s. Given $b_t, t = 1, \dots, T$, C' could be taken as $\max_t b_t$ for example.

We can now state our main result.

THEOREM 3.3.3 *Consider the DSFM (3.6) for $R \geq 3, T, J \geq 1$ and $\mathcal{T} = \{1, \dots, T\}$. Let also Assumption 3.3.3 be satisfied for the random vectors $\varepsilon_1, \dots, \varepsilon_T$ and $\Psi_t \Psi_t^\top / J = I_K, \sum_{t=1}^T U_t^\top U_t / R = 1$, and $M(\beta^*) \leq s$. Furthermore let κ be defined as in Assumption 3.3.1 and ϕ_{\max} is the maximum eigenvalue of the matrix $\sum_{t=1}^T U_t U_t^\top$. Let*

$$\lambda = \frac{C'}{\sqrt{T}} + \sqrt{\frac{\mathcal{X}^*(\mathcal{T}) \sum_t b_t^2}{(\log R)^{1-\delta'} T^2}}, \quad \delta' > 0.$$

Then with probability at least $p(1 - R^{-\delta'})$, for any solution $\hat{\beta}$ of (3.7) we have:

$$(JT)^{-1} \sum_{t=1}^T \left\| \Psi_t^\top (\hat{\beta} - \beta^*) U_t \right\|^2 \leq 16 \left(C' + \sqrt{\frac{\mathcal{X}^*(\mathcal{T}) \sum_t b_t^2}{(\log R)^{1-\delta'} T}} \right)^2 s / \kappa^2$$

$$T^{-1/2} \left\| \hat{\beta} - \beta^* \right\|_{2,1} \leq 16 \left(C' + \sqrt{\frac{\mathcal{X}^*(\mathcal{T}) \sum_t b_t^2}{(\log R)^{1-\delta'} T}} \right) s / \kappa^2$$

and

$$M(\hat{\beta}) \leq 64 \phi_{\max}^2 s / \kappa^2$$

Not surprisingly, this theorem tells that the bounds get larger when the dependence level, i.e. $\mathcal{X}^*(\mathcal{T})$ increases, i.e. the bound is minimized when $\mathcal{X}^*(\mathcal{T}) = 1$.

3.4 Generalized Dynamic Semiparametric Factor Model

The original model (3.6) assumes that there is no stochastic evolution in time. To this end, we consider the following extension of (3.4) and (3.6):

$$\begin{aligned} Z_{tl} &= \sum_{r=1}^R \gamma_{rl} u_r(t) \\ Y_t^\top &= (Z_{0,t}^\top + U_t^\top \Gamma) A \Psi_t + \varepsilon'_t = U_t^\top \Gamma A \Psi_t + (Z_{0,t}^\top A \Psi_t + \varepsilon'_t), \end{aligned} \quad (3.17)$$

with an unobservable L -dimensional random process $Z_{0,t}$ with $E(Z_{0,t}|X_t) = 0$ and i.i.d. assumption on ε'_t . We call (3.17) a generalized dynamic semiparametric factor model (GDSFM). If we concentrate on prediction, the trend represented by $U_t^\top \Gamma$ is enough. However, if we are interested in the stochasticity or dynamics of the original high dimensional time series, $Z_{0,t}$ comes into play, e.g. for pricing weather derivatives and various other financial engineering examples. The estimation procedure is now divided into 2 steps:

- For the model $Y_t^\top = U_t^\top \Gamma A \Psi_t + (Z_{0,t}^\top A \Psi_t + \varepsilon'_t)$, treat $Z_{0,t}^\top A \Psi_t + \varepsilon'_t$ as the ε_t in (3.6) and find the best parametric approximation according to the estimation procedure described in Subsection 3.2.3 to get the deterministic trend $U_t^\top \Gamma$.
- Based on $\hat{Y}_t^\top \stackrel{\text{def}}{=} Y_t^\top - U_t^\top \hat{\beta} \Psi_t$, \hat{A} and Ψ_t , use the ordinary least square method to obtain the estimated random process $\hat{Z}_{0,t}$.

As we could see from step one here, since ε_t in (3.6) involves $Z_{0,t}^\top A \Psi_t + \varepsilon'_t$, where $Z_{0,t}$ is a random process inhering dependence structure, Theorem 3.3.3 shows its necessity again. In the second step, $Z_{0,t}$ is estimated based on $\hat{\beta}$ instead of β^* , we need to show the influence of this plug-in estimate is negligible. Our first result this section relies on the following assumptions, which are similar to Assumptions (A1-8) in Park et al. (2009).

ASSUMPTION 3.4.1 *4.1.1 The variables $X_{1,1}, \dots, X_{T,J}$, $\varepsilon'_{1,1}, \dots, \varepsilon'_{T,J}$, and $Z_{0,1}, \dots, Z_{0,T}$ are independent.*

4.1.2 For $t = 1, \dots, T$ the variables $X_{t,1}, \dots, X_{t,J}$ are identically distributed, have support $[0, 1]^d$ and a density f_t that is bounded from below and above on $[0, 1]^d$, uniformly over $t = 1, \dots, T$.

4.1.3 We assume that $E\varepsilon'_{t,j} = 0$ for $1 \leq t \leq T, 1 \leq j \leq J$, and for $c > 0$ small enough $\sup_{1 \leq t \leq T, 1 \leq j \leq J} E \exp\{c(\varepsilon'_{t,j})^2\} < \infty$.

4.1.4 The vector of functions $m = (m_1, \dots, m_L)^\top$ can be approximated by Ψ_k , i.e.

$$\delta_K \stackrel{\text{def}}{=} \sup_{x \in [0,1]^d} \inf_{A \in \mathbb{R}^{L \times K}} \|m(x) - A \Psi(x)\| \rightarrow 0$$

as $K \rightarrow \infty$. We denote A that fulfills $\sup_{x \in [0,1]^d} \|m(x) - A \Psi(x)\| \leq 2\delta_K$ by A^ .*

3 Generalized Dynamic Semiparametric Factor Models

4.1.5 There exist constants $0 < C_L < C_U < \infty$ such that all eigenvalues of the matrix $T^{-1} \sum_{t=1}^T Z_{0t} Z_{0t}^\top$ lie in the interval $[C_L, C_U]$ with probability tending to one.

4.1.6 The minimization (3.7) runs over all values β with

$$\sup_{x \in [0,1]^d} \max_{1 \leq t \leq T} \|Z_{0,t}^\top A \Psi(x)\| \leq M_T,$$

where the constant M_T fulfils $\max_{1 \leq t \leq T} \|Z_{0,t}\| \leq M_T/C_m$ (with probability tending to one) for a constant C_m such that $\sup_{x \in [0,1]^d} \|m(x)\| < C_m$.

4.1.7 It holds that $\rho^2 = (K + T)M_T^2 \log(JTM_T)/(JT) \rightarrow 0$. The dimension L is fixed.

Assumption (4.1.6) and the additional bound M_T in the minimization is introduced for purely technical reasons.

THEOREM 3.4.1 Suppose that model (3.17), all assumptions in Theorem 3.3.3 and Assumption 3.4.1 hold. Then we have

$$\frac{1}{T} \sum_{1 \leq t \leq T} \left\| \widehat{Z}_{0,t}^\top \widehat{A} - Z_{0,t}^\top A^* \right\|^2 = \mathcal{O}_P(\rho^2 + \delta_K^2). \quad (3.18)$$

In the following we discuss how a statistical analysis differs if the inference of stochasticity on $Z_{0,t}$ is based on $\widehat{Z}_{0,t}$ (note that the trend $U_t^\top \Gamma$ is deterministic) instead of using (the unobserved) process $Z_{0,t}$. We will show that the differences are asymptotically negligible (up to an orthogonal transformation). This is the content of the following theorem, where we consider estimators of autocovariances and show that these estimators differ only by second order terms. This asymptotic equivalence carries over to classical estimation and testing procedures in the framework of fitting a vector autoregressive model. For the statement of the theorem we need the following assumptions, which are similar to Assumptions (A9-11) in Park et al. (2009):

ASSUMPTION 3.4.2 4.2.1 $Z_{0,t}$ is a strictly stationary sequence with $E(Z_{0,t}) = 0$, $E(\|Z_{0,t}\|^\gamma) < \infty$ for some $\gamma > 2$. It is strongly mixing with $\sum_{i=1}^\infty \alpha(i)^{(\gamma-2)/\gamma} < \infty$. The matrix $E Z_{0,t} Z_{0,t}^\top$ has full rank. The process $Z_{0,t}$ is independent of $X_{11}, \dots, X_{TJ}, \varepsilon'_{11}, \dots, \varepsilon'_{TJ}$.

4.2.2 It holds that $[\log(KT)^2 \{(KM_T/J)^{1/2} + T^{1/2} M_T^4 J^{-2} + K^{3/2} J^{-1} + K^{4/3} J^{-2/3} T^{-1/6}\} + 1] T^{1/2} (\rho^2 + \delta_K^2) = \mathcal{O}(\rho^2 + \delta_K^2)$

Assumption (4.2.2) poses very weak conditions on the growth of J, K, T . Suppose, for example, that M_T is of logarithmic order and that K is of order $(JT)^{1/5}$ so that the variance and the bias are balanced for twice differentiable functions. In this setting, (4.2.1) only requires that T/J^2 times a logarithmic factor converges to zero.

Furthermore, please note that the minimization problem (3.7) has only a unique solution up to β , but not to Γ, A . If $(\widehat{Z}_{0,t}, \widehat{A})$ is a minimizer, then also $(B^\top \widehat{Z}_{0,t}, B^{-1} A)$ is a minimizer, where B is an arbitrary invertible matrix. In particular, with the choice

$B = (\sum_{t=1}^T Z_{0,t} \hat{Z}_{0,t})^{-1} \sum_{t=1}^T Z_{0,t} Z_{0,t}^\top$, we get for $\tilde{Z}_{0,t} \stackrel{\text{def}}{=} B^\top \hat{Z}_{0,t}$ and $\tilde{A} \stackrel{\text{def}}{=} B^{-1} A$ that $\sum_{t=1}^T Z_{0,t} (\tilde{Z}_{0,t} - Z_{0,t})^\top = 0$. Without loss of generality, we may assume $T^{-1} \sum_{s=1}^T \hat{Z}_{0,s} = T^{-1} \sum_{s=1}^T Z_{0,s} = 0$. Additionally define

$$\begin{aligned}\tilde{Z}_{n,t} &= (T^{-1} \sum_{s=1}^T \tilde{Z}_{0,s} \tilde{Z}_{0,s}^\top)^{-1/2} \tilde{Z}_{0,t} \\ Z_{n,t} &= (T^{-1} \sum_{s=1}^T Z_{0,s} Z_{0,s}^\top)^{-1/2} Z_{0,t}.\end{aligned}$$

THEOREM 3.4.2 *Suppose that model (3.17) holds. Besides all assumptions in Theorem 3.3.3, let also Assumption 3.4.1-3.4.2 be satisfied. Then there exists a random matrix B such that for $h \geq 0$*

$$T^{-1} \sum_{t=\max[1, -h+1]}^{\min[T, T-h]} \tilde{Z}_{0,t} (\tilde{Z}_{0,t+h} - \tilde{Z}_{0,t})^\top - Z_{0,t} (Z_{0,t+h} - Z_{0,t})^\top = \mathcal{O}_P(T^{-1/2})$$

and

$$T^{-1} \sum_{t=\max[1, -h+1]}^{\min[T, T-h]} \tilde{Z}_{n,t} \tilde{Z}_{n,t+h}^\top - Z_{n,t} Z_{n,t+h}^\top = \mathcal{O}_P(T^{-1/2}).$$

3.5 Simulation Study

We present three simulations which investigate how the spread of the sparsity level $M(\beta^*)$, the number of initial time basis R and the dependence level of the error terms affect the performance. In the first example, we show how changing the values of $M(\beta^*)$ result in changing the two measures of estimation error in light of Theorem 3.3.1:

$$\begin{aligned}L_{par} &= 1 - \frac{\sum_{r=1}^R \|\hat{\beta}_r - \beta_r\|_\infty}{\sum_{r=1}^R \|\beta_r\|_\infty} \\ L_{pre} &= 1 - \frac{\sum_{r=1}^R \|\Psi_t^\top (\hat{\beta}_r - \beta_r) U_t\|_\infty}{\sum_{r=1}^R \|\Psi_t^\top \beta_r U_t\|_\infty}\end{aligned}$$

All codes were done in Matlab and are available on the author's homepage or www.quantlet.com. We applied the above algorithm (3.8), (3.9) to the following simulated data. We generate random $\beta_1, \dots, \beta_{179} \in \mathbb{R}^5$ such that all coordinates are independent and consider an initial model with the parameters such that $\beta_{rk} \sim N\{0, \exp(-2k/5)\}$, $r = 1, \dots, 179, k = 1, \dots, 5$. We randomly pick $179 - M(\beta^*)$ β_r s from $\beta_1, \dots, \beta_{179}$ and assign them to be $0 \in \mathbb{R}^5$. We choose the same time basis as in Table 3.4. For the space part, inspired by Park et al. (2009), we considered $d = 2, L = 3$

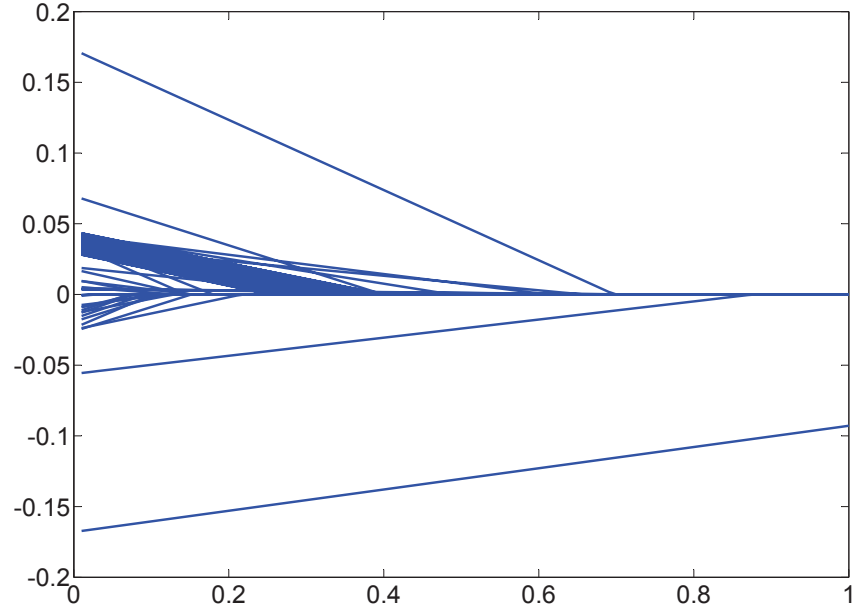


Figure 3.5: An illustration plot about how group Lasso penalty shrinks the coefficients.

and the following tuple of 2-dimensional functions:

$$\begin{aligned} m_0(x_1, x_2) &= 1, & m_1(x_1, x_2) &= 3.46(x_1 - .5), \\ m_2(x_1, x_2) &= 9.45 \left\{ (x_1 - .5)^2 + (x_2 - .5)^2 \right\} - 1.6, \\ m_3(x_1, x_2) &= 1.41 \sin(2\pi x_2). \end{aligned}$$

The coefficients in these functions were chosen so that m_1, m_2, m_3 are close to orthogonal. The design points $X_{t,j}$ were independently generated from a uniform distribution on the unit square. We generate $Y_t^\top = U_t^\top \beta^\top \Psi_t + \varepsilon_t, t = 1, \dots, 19345$ where ε_t is drawn as i.i.d. $N(0, 0.05)$.

The convergence of the algorithm presented in (3.10) is usually achieved up to 5 iterations. Figure 3.5 is an illustration plot about how the group Lasso penalty shrinks the coefficients.

With 250 repetitions, Table 3.1 displays different L_{par} and L_{pre} s w.r.t. different sparsity levels. Our theoretical results in the previous sections suggest that when $M(\beta^*)$ (s) is small, L_{par} and L_{pre} will be large, which is confirmed by the simulation results.

	$M(\beta^*) = 100$	$M(\beta^*) = 50$	$M(\beta^*) = 20$
L_{par}	0.870	0.918	0.931
L_{pre}	0.710	0.835	0.859

Table 3.1: L_{par} and L_{pre} w.r.t. different sparsity levels.

The second experiment compares how L_{par} and L_{pre} react to changing the numbers of initial time basis R , for $M(\beta^*) = 50$, if we additionally include the quartic term in the orthogonal polynomial and double the number of Fourier series, $R = 53 \cdot 4 + 40 = 252$ and if we remove the cubic term in the orthogonal polynomial and half the number of Fourier series, $R = 53 \cdot 2 + 10 = 116$. The L_{par} and L_{pre} are presented in Table 3.2.

	$R = 116$	$R = 179$	$R = 252$
L_{par}	0.879	0.918	0.920
L_{pre}	0.695	0.835	0.841

Table 3.2: L_{par} and L_{pre} w.r.t. different number of initially involved time basis.

As we could see, when R increases, L_{par} and L_{pre} increase. This indicates us that in practice we need take a relatively large R value, i.e. involve as many as possible time basis.

The third experiment compares how L_{par} and L_{pre} are sensitive to the dependence level of the error items. We generated ε_t from a centered VAR(1) process $\varepsilon_t = \mathcal{R}\varepsilon_{t-1} + U_t$, where U_t is $N_3(0, \Sigma_U)$ random vector, the rows of \mathcal{R} from the top equal $(0.95, -0.2, 0)$, $(0, 0.8, 0.1)$, $(0.1, 0.06)$, and $\Sigma_U = 10^{-4}I_3$. We choose $M(\beta^*) = 50$, $R = 179$ as before. Besides the VAR(1) process indicated before, we also tried the VAR(2) to generate ε_t . Table 3.3 displays the result, where we use VAR(0) to denote the independent case. The performance decreases when the error terms are more dependent, which is consistent with Theorem 3.3.3.

	VAR(0)	VAR(1)	VAR(2)
L_{par}	0.918	0.854	0.783
L_{pre}	0.835	0.774	0.712

Table 3.3: L_{par} and L_{pre} w.r.t. different levels of dependence of ε_t .

For more Monte Carlo experiments concerning Theorem 3.4.2, we refer to Park et al. (2009).

3.6 Weather, Neuro-economics and IVS

This section presents three applications to the temperature, fMRI and IVS analysis. First, we fit the model to the daily temperature observations by Climatic Data Center (CDC), China Meteorological Administration (CMA), as introduced in Figure 3.1. To capture the upward trend, seasonal and "large period" effects, for time basis, similar to Racsko et al. (1991), Parton and Logan (1981) and Hedin (1991), we propose the following initial choice of time basis (rescaling factors omitted) in Table 3.4.

For the space basis, consider the eigenvalues of the smoothed (with the usual optimal bandwidth for local polynomial regression) covariance operator (Figure 3.6) and also the climate types of China (Figure 3.7), the number of space basis $K = 5$ seems to be

3 Generalized Dynamic Semiparametric Factor Models

	Factors	Factors	
Trend (Year by Year)	1	Large	$\sin 2\pi t / (365 \cdot 10)$
	t	Period	$\cos 2\pi t / (365 \cdot 10)$
	$3t^2 - 1$		$\sin 4\pi t / (365 \cdot 10)$
Seasonal Effect	$\sin 2\pi t / 365$	$\cos 4\pi t / (365 \cdot 10)$	$\sin 6\pi t / (365 \cdot 10)$
	$\cos 2\pi t / 365$		$\sin 6\pi t / (365 \cdot 10)$
	\dots		\dots
	$\cos 20\pi t / 365$		$\cos 20\pi t / (365 \cdot 10)$

Table 3.4: Initial choice of $53 \cdot 3 + 20 = 179$ time basis.

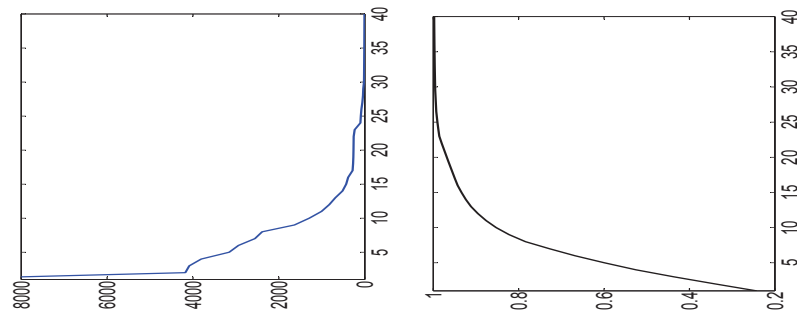


Figure 3.6: Distribution of the eigenvalues and the relative proportion of variance explained by the first K basis.

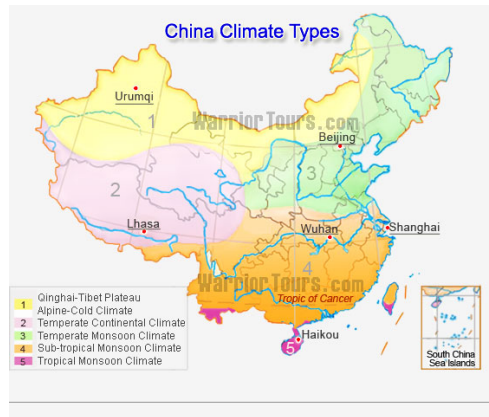


Figure 3.7: China Climate Types

satisfactory although $K = 10$ is needed to pass the “90%” rule. Please note that it is significantly smaller than the number of terms of a series estimator. Figure 3.8 displays the estimated coefficients of the 5 factors with respect to the $54 \cdot 3$ yearly polynomial time basis under the optimal choice of λ . The coefficients of constant, linear and quadratic terms are displayed as solid, dashed and dotted lines correspondingly. As one may see, the fact that most of the coefficients are nonnegative (especially for $k = 1$) shows strong evidence of global warming effect (especially with a quadratic upward trend) in China during the past 50 years. In a climatological context this has also been observed by Karl et al. (1991), while the global climate change has been recently summarized by Gleick et al. (2010). The high estimates over the second half of 1960s are due to the high temperatures then in China (Figure 3.1). The pattern that all the coefficients display an upward trend further indicates the stronger and stronger warming effect. The coefficients estimates of the 20 Fourier series time basis corresponding to the optimal λ are displayed in Table 3.5. It clearly indicates the 10-year period effect which, as some meteorologists claimed, are related to the solar activity. Figure 3.9 displays the extracted trends based on $U_t^\top \hat{\beta}$, where the five lines correspond to the five factors. The characters of this kind of nonstationary time series further indicate that the autoregressive model may not be a proper tool to capture them. Firstly, since there exists the “stronger and stronger global warming” effect, if we use AR model, the constant, linear and quadratic coefficients should be time variant (increasing). Secondly, the existence of “large period” effect also poses the problem of lag or frequency selections there. Both of these actually introduce bigger technical challenges.

Basis	Estimates				
$\sin 2\pi t/365$	−0.1777	0.0076	0.0177	−0.0136	0.0084
$\cos 2\pi t/365$	−0.6081	0.0126	0.0366	−0.0369	0.0114
$\sin 4\pi t/365$	0.0000	0.0000	0.0000	0.0000	0.0000
$\cos 4\pi t/365$	−0.0145	0.0028	0.0021	−0.0022	0.0029
...	0.0000	...			
$\cos 20\pi t/365$	0.0000	...			
$\sin 2\pi t/(365 \cdot 10)$	0.0025	−0.0006	0.0009	−0.0008	−0.0001
$\cos 2\pi t/(365 \cdot 10)$	0.0000	...			
...	0.0000	...			
$\cos 20\pi t/(365 \cdot 10)$	0.0000	...			

Table 3.5: Estimated coefficients of the 5 factors w.r.t. the 20 Fourier series time basis.

Since the eigenvalues of $\hat{\beta}\hat{\beta}^\top$ are $(0.4683, 0.0106, 0.0068, 0.0040, 0.0007, 0.0000, \dots)$, we choose $L = 5$ and estimated the remaining 5-dimensional random process $\hat{Z}_{0,t}$, e.g. $\hat{Z}_{0,t,1}$ as displayed in Figure 3.10 ($\hat{Z}_{0,t,2} - \hat{Z}_{0,t,5}$ are omitted due to the limited space here). The expectation of the random process is close to zero, which indicates our detrending using the group Lasso type technique works well. The residual multi-dimensional random process could be further modeled by multivariate time series techniques. For example, if we use VAR(1) process $\hat{Z}_{0,t} = \mathcal{R}\hat{Z}_{0,t-1} + \varepsilon_{0,t}$, where $\varepsilon_{0,t}$ is a random vector, the estimated

3 Generalized Dynamic Semiparametric Factor Models

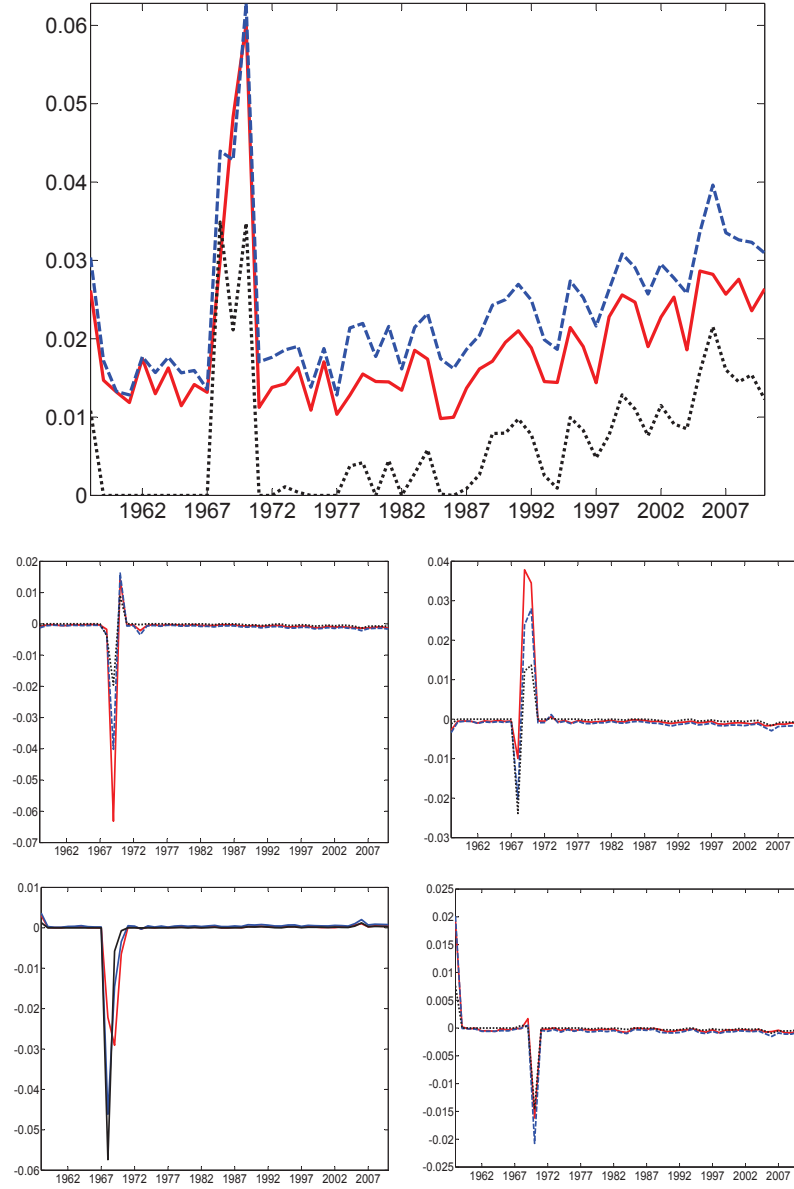


Figure 3.8: Estimated coefficients of the $54 \cdot 3$ yearly polynomial time basis w.r.t. $k = 1, \dots, 5$ from up to down and left to right.

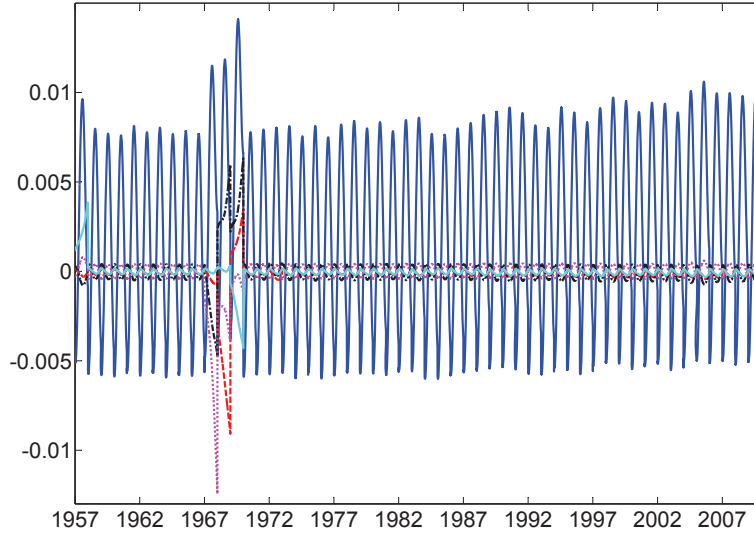


Figure 3.9: Extracted trends based on $U_t^\top \hat{\beta}$.

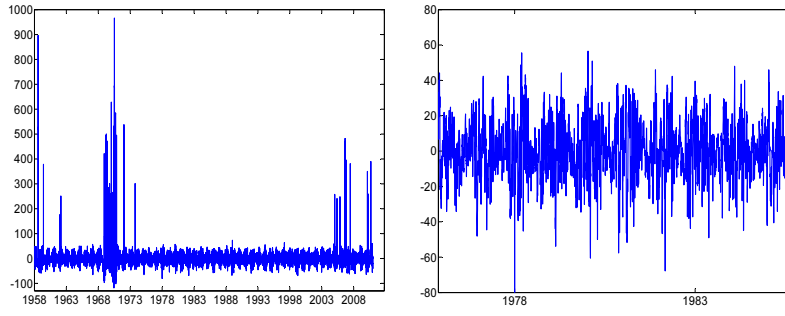


Figure 3.10: Estimated Stochastic Process $\hat{Z}_{0,t,1}$ and a 10 year zoom.

3 Generalized Dynamic Semiparametric Factor Models

coefficient matrix is:

$$\begin{pmatrix} 0.9732 & -0.0135 & -0.0002 & -0.0006 & -0.0002 \\ 0.0127 & 0.1766 & -0.1824 & -0.0682 & -0.0009 \\ 0.0358 & -0.2867 & 0.4493 & -0.1138 & 0.0053 \\ -0.0001 & -0.1967 & -0.1962 & 0.8010 & -0.0052 \\ 0.0790 & 0.0492 & 0.0690 & -0.0225 & 0.8418 \end{pmatrix}.$$

In comparison with the existing temperature modeling or weather derivatives pricing techniques, e.g. Benth and Benth (2005), we have the following advantages. Firstly, based on the high dimensional time series data, we offer integrated analysis considering space (high dimensionality) and time (dynamics) parts simultaneously, while forecasting at places different from the existing weather stations is also possible since the space basis are actually functions of the geographical location information. Secondly, we extract the trend more clearly. Thirdly, we provide the theoretical justification for further inferential analysis of $\hat{Z}_{0,t}$ instead of $Z_{0,t}$. However, if we have a closer look at the enlarged estimated stochastic process in Figure 3.10, we find that the volatility of the random process also has a seasonality, which is actually due to the fact that the variance of the noise (temperature, fMRI etc.) scale linearly with the expectation of the measurements. This motivates to consider (3.6) under heteroscedasticity (Poisson - like model) as follows:

$$Y_t^\top = U_t^\top \Gamma A \Psi_t + \varepsilon_t, \quad \text{Cov}(\varepsilon_t) = \text{diag}(|U_t^\top \Gamma A \Psi_t|),$$

which will be presented in a separate paper.

As a second application of the model, we consider a microeconomic experiment based on fitting an fMRI data set. Here we used a novel investment decision task that uses streams of (past) returns as stimuli to the exercised subjects, where the flowchart of the experiment is presented in Figure 3.11 (left), and obtain a series of three-dimensional images of the blood-oxygen-level-dependent (BOLD) fMRI signals. Our model helps to identify the corresponding brain's activation areas and to simplify the inference to the analysis of time propagation of a few number of factors (low-dimensional representation). Additionally we classify the risk attitudes of different subjects based on the coefficients of time basis, which performed quite well compared to the classic risky decision making model (risk-return model) which is based on the subjects' answers directly, where the risk attitude can be measured as value reduction in Euro for maximum risk (the case when the subjective perceived risk = 100), as described in Mohr et al. (2010). All subjects were classified as risk averse indicated by a positive risk weight as shown in Figure 3.11 (right). However, for six subjects the risk attitude was quite low (risk weight < 5, colored with blue) resulting in only a small influence of risk on value. For the experimental procedure and the fMRI data description, we refer to Myšičková et al. (2010).

Since we are analyzing multi subjects $1 \leq i \leq I$ here, we obtain a panel version of the

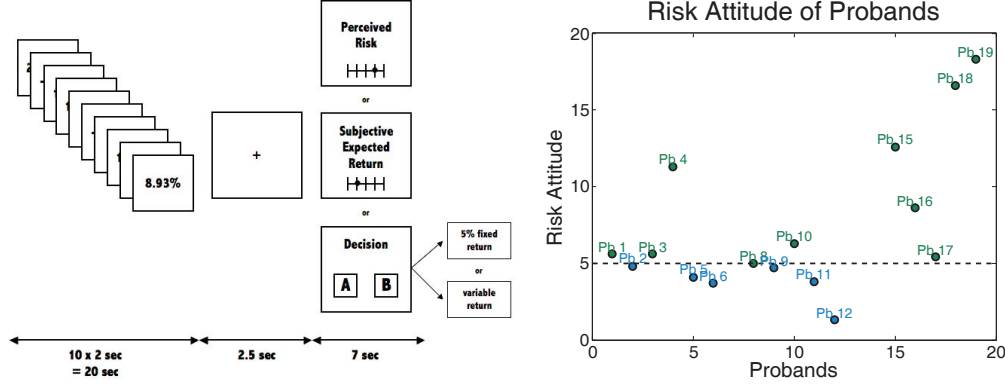


Figure 3.11: Flowchart of the experiment (left) “Returns Pause Decision” and risk attitudes of 16 subjects (right). Subjects with risk attitude < 5 are colored blue, otherwise green.

original model (3.6) to

$$Y_{t,j}^i = \sum_{l=1}^L (\alpha_{t,l}^i + U_t^\top \Gamma_l^i) m_l(X_{t,j}) + \varepsilon_{t,j}, \quad 1 \leq j \leq J_t, \quad 1 \leq t \leq T,$$

where the fixed effect $\alpha_{t,l}^i$ is the individual effect on function m_l for subject i at time point t . For identification purpose, we assume

$\sum_{i=1}^I \sum_{l=1}^L \alpha_{t,l}^i m_l(X_{t,j}) = 0$. Please notice that assuming different subjects have the same basis function in space m_l makes sense here since the basis function is used to detect which part of the brain is activated for risky decisions, which should be homogeneous for human beings. Thus for this panel data, we have:

$$\bar{Y}_{t,j} = \sum_{l=1}^L (U_t^\top \bar{\Gamma}_l) m_l(X_{t,j}) + \varepsilon_{t,j}, \quad 1 \leq j \leq J,$$

and our 2-step estimation procedure is as follows:

- 1 Take the average of $Y_{t,j}^i$ across different subjects i , and estimate the common basis function in space m_l as in the original approach.
- 2 Given the common m_l , for different subjects i , estimate their specific factors in time $Z_{t,l}^i$.

$$Y_{t,j}^i = \sum_{l=1}^L U_t^\top \Gamma_l^i \bar{m}_l(X_{t,j}) + \varepsilon_{t,j}^i$$

Since most of the technical details have been illustrated in the previous application, it is skipped here, while the differences will be emphasized. Since the significantly larger

dimension $J = 76176$ is observed here, computing eigenvalues of a 76176×76176 matrix will encounter significant numerical difficulties. By using the fact that cc^\top has the same eigenvalues as $c^\top c$ (where c is a $J \times T$ matrix), we only need to compute eigenvalues of a 722×722 matrix. If we additionally take the average of every 10 $Y_{t,j}^i$ s over t , we only need compute eigenvalues from a 73×73 matrix.

The third factor loading function \hat{m}_3 shown in Figure 3.12 could be identified as the Ventromedial prefrontal cortex (VMPFC) located in the bottom frontal part in the brain, which is the center for utility and conform herewith with our experiment (it is why it is presented here). The other functions m_l , which also represent exactly those brain regions which we have expected to be involved during the experiment, are presented in Myšičková et al. (2010).

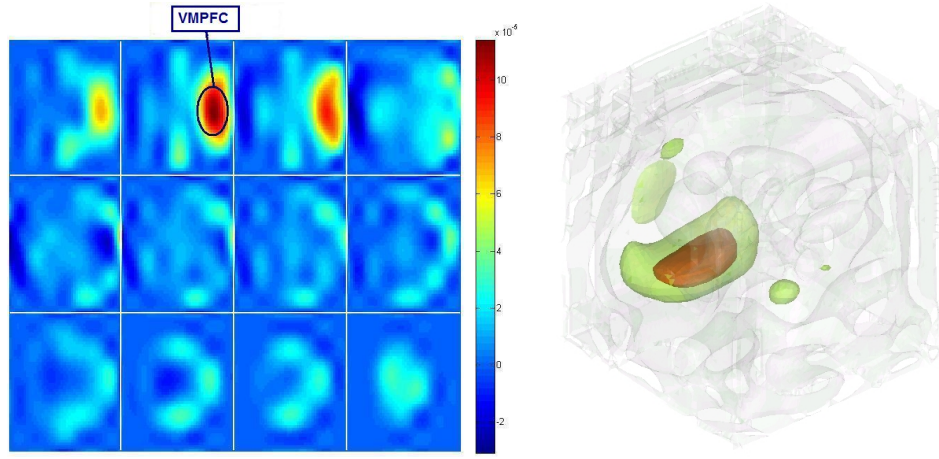


Figure 3.12: Estimated function \hat{m}_3 shown in 12 axial slices (left) and as a 3D-plot in a posterior view (right) with highlighted Ventromedial prefrontal cortex (VMPFC).

We use cubic orthogonal polynomials as time basis. Figure 3.13 displays the response curve (to the “decision phase” stimuli) $U_t^\top \hat{\Gamma}_2^i$ for different subjects. Based on the estimated factors for different individuals, we could further develop a classification method which can predict the risk aversion only based on the measured fMRI signals. Observing that different probands’ response curves have different patterns and their corresponding $\hat{Z}_{0,t}$ have different volatilities, for this purpose we use the estimated coefficients Γ_3^i since it correspond to the brain activity of the VMPFC, which is linked with utility. To provide the classification analysis, we apply Support Vector Machines (SVM), which is a widely used nonlinear method based on statistical learning theory. For the learning step, strongly risk averse subjects were labeled by -1 and weakly risk averse subjects by 1 . Then, we applied the leave-one-out method to first train and then estimate the classification rate of the SVM. The classification rates are 85% for strongly risk averse and 60% for weakly risk averse individuals. More importantly, these rates hold for a wide range of prior parameters: the radial basis coefficient r ($0.25 - 0.35$) and the capacity C

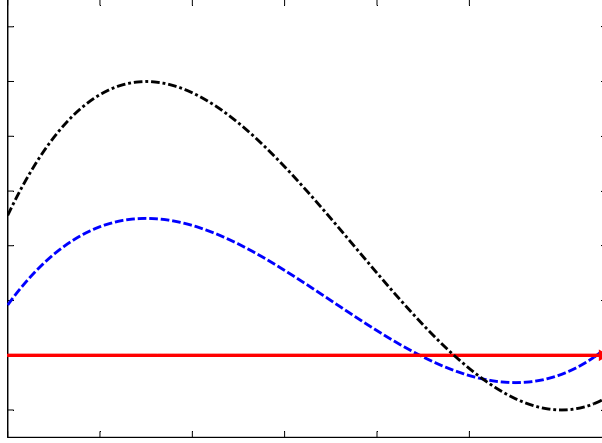


Figure 3.13: Response curves (to stimuli) $U_t^\top \hat{\Gamma}_2^i$ for probands $i = 9$ (blue dashed) & $i = 19$ (black dash dotted) with periodic cubic polynomial as time basis.

(20 – 90).

MEAN		Estimated	
Data	Strongly	0.85	0.14
	Weakly	0.59	0.40

Table 3.6: Classification rates of the SVM method using median(left) and mean (right) of volatilities of $\Delta \hat{Z}_{t,2}$.

In the analysis of IVS data, deterministic trends are not present, and do not make sense from a non arbitrage point of view. We may therefore assume stationarity. The first detrending step is therefore omitted, alternatively, we could still use the dynamic semiparametric factor modeling approach proposed by Park et al. (2009) except for a different space basis. To this end, due to the limited space here, we only present the new space basis of the implied volatility surface (IVS) application in Figure 3.14 (left). We see that the “smile” effect is captured very well. The corresponding estimated time series of factors $\hat{Z}_{t,1}$, $\hat{Z}_{t,2}$ (stationary) are presented in Figure 3.14 (right).

3.7 Appendix

Here we collect one auxiliary result which is used in the proof of Lemma 3.3.1.

LEMMA 3.7.1 *For any $I \times J$ matrix A and any $J \times K$ matrix B , we have $\|AB\|_{2,1} \leq \|A\|_{2,1}\|B\|_{2,1}$.*

3 Generalized Dynamic Semiparametric Factor Models

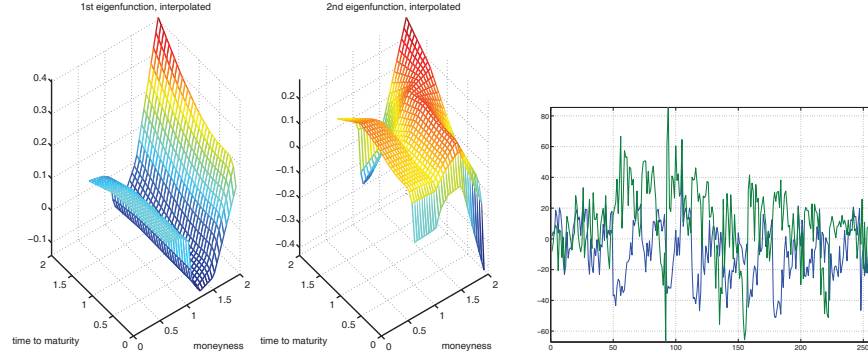


Figure 3.14: Space basis using the FPCA approach for IVS modeling and the estimated time series of factors $\hat{Z}_{t,1}$, $\hat{Z}_{t,2}$.

Proof With Cauchy Schwartz inequality it is not hard to derive:

$$\begin{aligned}
 \|AB\|_{2,1} &= \sum_{i=1}^I \sqrt{\sum_{k=1}^K \left(\sum_{j=1}^J a_{ij} b_{jk} \right)^2} \\
 &\leq \sum_{i=1}^I \sqrt{\sum_{k=1}^K \left(\sum_{j=1}^J a_{ij}^2 \sum_{j=1}^J b_{jk}^2 \right)} \\
 &\leq \left(\sum_{i=1}^I \sqrt{\sum_{j=1}^J a_{ij}^2} \right) \left(\sum_{k=1}^K \sqrt{\sum_{j=1}^J b_{jk}^2} \right) \\
 &= \|A\|_{2,1} \|B\|_{2,1}
 \end{aligned}$$

□

Proof of Lemma 3.3.1 The proof is in a similar spirit of the one of Lemma 3.1 in Lounici et al. (2009). By the definition of $\hat{\beta}$ as a minimizer of (3.7), for $\forall \beta$ we have

$$\begin{aligned}
 (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top \hat{\beta} U_t - Y_t\|^2 + 2\lambda \sum_{r=1}^R \|\hat{\beta}_r\| \\
 \leq (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top \beta U_t - Y_t\|^2 + 2\lambda \sum_{r=1}^R \|\beta_r\|,
 \end{aligned} \tag{3.19}$$

which, using $Y_t = \Psi_t^\top \beta^* U_t + \varepsilon_t$, is equivalent to

$$\begin{aligned} (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\hat{\beta} - \beta^*) U_t\|^2 &\leq (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\beta - \beta^*) U_t\|^2 \\ &\quad + 2(JT)^{-1} \sum_{t=1}^T \varepsilon_t^\top \Psi_t^\top (\hat{\beta} - \beta) U_t + 2\lambda \sum_{r=1}^R (\|\beta_r\| - \|\hat{\beta}_r\|). \end{aligned} \quad (3.20)$$

By Hölder's inequality, we have that

$$\sum_{t=1}^T \varepsilon_t^\top \Psi_t^\top (\hat{\beta} - \beta) U_t \leq \left\| \sum_{t=1}^T \Psi_t \varepsilon_t U_t^\top \right\|_{2,\infty} \|\hat{\beta} - \beta\|_{2,1} \quad (3.21)$$

where $\left\| \sum_{t=1}^T \Psi_t \varepsilon_t U_t^\top \right\|_{2,\infty} = \max_{1 \leq r \leq R} \sqrt{\sum_{t=1}^T \sum_{k=1}^K (\sum_{j=1}^J \Psi_{tkj}^\top \varepsilon_{tj} U_{tr})^2}$.

Consider the random event

$$\mathcal{A} = \left\{ 2(JT)^{-1} \left\| \sum_{t=1}^T \Psi_t \varepsilon_t U_t^\top \right\|_{2,\infty} \leq \lambda \right\}. \quad (3.22)$$

Since $\Psi_t \Psi_t^\top / J = I_K$ and $\sum_{t=1}^T U_t^\top U_t / R = 1$, the random variables $V_{tr} = (\sqrt{J}\sigma)^{-1/2} \sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr}$, $t = 1, \dots, T$, are *i.i.d.* standard Gaussian.

Using this fact, we can write, for any $r = 1, \dots, R$, and $\lambda = 2\sigma/\sqrt{JT} (1 + A \log R/\sqrt{T})^{1/2}$,

$$\begin{aligned} \mathbb{P} \left\{ \sum_{t=1}^T \sum_{k=1}^K \left(\sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \geq \lambda^2 (JT)^2 / 4 \right\} &= \mathbb{P} \left\{ \mathcal{X}_T^2 \geq \lambda^2 JT^2 / (4\sigma^2) \right\} \\ &= \mathbb{P} \left(\mathcal{X}_T^2 \geq T + A\sqrt{T} \log R \right) \end{aligned}$$

where \mathcal{X}_T^2 is a chi-square random variable with T degrees of freedom. By the tail property of \mathcal{X}_T^2 distribution (Lemma A.1 of Lounici et al. (2009)), and the fact that $A > 8$ we get:

$$\mathbb{P}(\mathcal{A}^c) \leq R \exp\{-A \log R / 8 \min(\sqrt{T}, A \log R)\} \leq R^{1-q}$$

3 Generalized Dynamic Semiparametric Factor Models

with $q = \min(A \log R, \sqrt{T})$. It follows from (3.20) and (3.21) that, on the event \mathcal{A} :

$$\begin{aligned}
& (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\hat{\beta} - \beta^*) U_t\|^2 + \lambda \sum_{r=1}^R \|\hat{\beta}_r - \beta_r\| \\
& \leq (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\beta - \beta^*) U_t\|^2 + 2\lambda \sum_{r=1}^R (\|\hat{\beta}_r - \beta_r\| + \|\beta_r\| - \|\hat{\beta}_r\|) \\
& \leq (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\beta - \beta^*) U_t\|^2 + 2\lambda \sum_{r \in \mathcal{R}(\beta)} (\|\hat{\beta}_r - \beta_r\| + \|\beta_r\| - \|\hat{\beta}_r\|) \quad (3.23) \\
& \quad + 2\lambda \sum_{r \in \mathcal{R}^c(\beta)} (\|\hat{\beta}_r - \beta_r\| + \|\beta_r\| - \|\hat{\beta}_r\|) \\
& \leq (JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\beta - \beta^*) U_t\|^2 + 4\lambda \sum_{r \in \mathcal{R}(\beta)} \|\hat{\beta}_r - \beta_r\|
\end{aligned}$$

which coincides with (3.11). To prove (3.12), we use (3.8) and (3.9) resulting in the inequality

$$(JT)^{-1} \max_{1 \leq r \leq R} \left\| \sum_{t=1}^T \{\Psi_t(Y_t - \Psi_t^\top \hat{\beta} U_t) U_t^\top\}_r \right\| \leq \lambda. \quad (3.24)$$

Then

$$\begin{aligned}
& (JT)^{-1} \left\| \sum_{t=1}^T \{\Psi_t \Psi_t^\top (\hat{\beta} - \beta^*) U_t U_t^\top\}_r \right\| \\
& \leq (JT)^{-1} \left\| \sum_{t=1}^T \{\Psi_t (\Psi_t^\top \hat{\beta} U_t - Y_t) U_t^\top\}_r \right\| + (JT)^{-1} \left\| \sum_{t=1}^T (\Psi_t \varepsilon_t U_t^\top)_r \right\| \quad (3.25)
\end{aligned}$$

where we have used $Y_t = \Psi_t^\top \beta^* U_t + \varepsilon_t$ and the triangle inequality. The derived bound (3.12) then follows by combining (3.25) with (3.24) and using the definition of the event \mathcal{A} . Finally, we prove (3.13). First, observe that,

$$\sum_{t=1}^T \Psi_t (Y_t - \Psi_t^\top \beta^* U_t) U_t^\top = \sum_{t=1}^T \Psi_t \Psi_t^\top (\hat{\beta} - \beta^*) U_t U_t^\top + \sum_{t=1}^T \Psi_t \varepsilon_t U_t^\top.$$

On the event \mathcal{A} , following from (3.8) and the triangle inequality, we have:

$$(JT)^{-1} \left\| \sum_{t=1}^T \{\Psi_t \Psi_t^\top (\hat{\beta} - \beta^*) U_t U_t^\top\}_r \right\| \geq \lambda/2, \text{ if } \hat{\beta}_r \neq 0.$$

The following arguments yields the bound (3.13) on the number of nonzero rows of $\widehat{\beta}_r^\top$:

$$\begin{aligned}
M(\widehat{\beta}) &\leq \frac{4}{\lambda^2(JT)^2} \sum_{r \in \mathcal{R}(\widehat{\beta})} \left\| \sum_{t=1}^T \{\Psi_t \Psi_t^\top (\widehat{\beta} - \beta^*) U_t U_t^\top\}_r \right\|^2 \\
&\leq \frac{4}{\lambda^2(JT)^2} \sum_{r=1}^R \left\| \sum_{t=1}^T \{\Psi_t \Psi_t^\top (\widehat{\beta} - \beta^*) U_t U_t^\top\}_r \right\|^2 \\
&= \frac{4}{\lambda^2 T^2} \left\| \sum_{t=1}^T \{J^{-1} \Psi_t \Psi_t^\top (\widehat{\beta} - \beta^*) U_t U_t^\top\} \right\|_{2,1}^2 \\
&\leq \frac{4}{\lambda^2 T^2} \|\widehat{\beta} - \beta^*\|_{2,1}^2 \left\| \sum_{t=1}^T U_t U_t^\top \right\|_{2,1}^2 \\
&\leq \frac{4\phi_{max}^2}{\lambda^2 T^2} \|\widehat{\beta} - \beta^*\|_{2,1}^2,
\end{aligned}$$

which follows from Lemma 3.7.1, $\Psi_t \Psi_t^\top / J = I_K$ and ϕ_{max} is the maximum eigenvalues of the matrix $\sum_{t=1}^T U_t U_t^\top$. \square

Proof of Theorem 3.3.1 We proceed similarly to the proof of Theorem 3.1 in Lounici et al. (2009) and Theorem 6.2 in Bickel et al. (2008). Let $\mathcal{R} = \mathcal{R}(\beta^*) = \{r : \beta_r^* \neq 0\}$

By inequality (3.11) in Lemma 3.3.1 with $\beta = \beta^*$ we have, on the event \mathcal{A} defined in (3.22):

$$\begin{aligned}
(JT)^{-1} \sum_{t=1}^T \|\Psi_t^\top (\widehat{\beta} - \beta^*) U_t\|^2 &\leq 4\lambda \sum_{r \in \mathcal{R}} \|\widehat{\beta}_r - \beta_r^*\| \\
&\leq 4\lambda \sqrt{s} \|(\widehat{\beta} - \beta^*)_{\mathcal{R}}\|
\end{aligned} \tag{3.26}$$

Moreover by the same inequality, on the event \mathcal{A} , we have $\sum_{r=1}^R \|\widehat{\beta}_r - \beta_r^*\| \leq 4 \sum_{r \in \mathcal{R}} \|\widehat{\beta}_r - \beta_r^*\|$, which implies that $\sum_{r \in \mathcal{R}^c} \|\widehat{\beta}_r - \beta_r^*\| \leq 3 \sum_{r \in \mathcal{R}} \|\widehat{\beta}_r - \beta_r^*\|$. Thus, by Assumption 3.3.1 with $\Delta = (\widehat{\beta} - \beta^*)$:

$$\|(\widehat{\beta} - \beta^*)_{\mathcal{R}}\| \leq \sum_{t=1}^T \|\Psi_t^\top (\widehat{\beta} - \beta^*) U_t\| / (\kappa \sqrt{J}). \tag{3.27}$$

Now (3.14) follows from (3.26) and (3.27). Inequality (3.15) follows by noting that

$$\sum_{r=1}^R \|\widehat{\beta}_r - \beta_r^*\| \leq 4 \sum_{r \in \mathcal{R}} \|\widehat{\beta}_r - \beta_r^*\| \leq 4\sqrt{s} \|(\widehat{\beta} - \beta^*)_{\mathcal{R}}\|$$

and then using (3.14). Inequality (3.16) follows from (3.13) and (3.14). \square

3 Generalized Dynamic Semiparametric Factor Models

Proof of Theorem 3.3.2 The proofs of this theorem are similar to the one of Theorem 3.3.1 up to a modification of the bound on $P(\mathcal{A}^c)$ in Lemma 3.3.1. We consider now the event

$$\mathcal{A} = \left\{ \max_{1 \leq r \leq R} \left\{ \sum_{t=1}^T \sum_{k=1}^K \left(\sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\}^{1/2} \leq \lambda JT \right\}.$$

The Markov inequality yields that

$$P(\mathcal{A}^c) \leq \sum_{t=1}^T E \left\{ \max_{1 \leq r \leq R} \left(\sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\} / (\lambda JT)^2.$$

Then we use Nemirovski's inequality, see Corollary 2.4 of Dümbgen et al. (2008)[p.5], with the random vectors

$$W_{tj} = \left(\sum_{k=1}^K \Psi_{tkj} \varepsilon_{tj} U_{t1} / J, \dots, \sum_{k=1}^K \Psi_{tkj} \varepsilon_{tj} U_{tR} / J \right) \in \mathbb{R}^R, \quad \forall j, \forall t.$$

We get that

$$P(\mathcal{A}^c) \leq \frac{2e \log R - e}{\lambda^2 JT} \sigma^2 (JT)^{-1} \sum_{t=1}^T \sum_{j=1}^J \left(\max_{1 \leq r \leq R} \left| \sum_{k=1}^K \Psi_{tkj} U_{tr} \right| \right)^2.$$

By the definition of λ in Theorem 3.3.2 and Assumption 3.3.2 we obtain

$$P(\mathcal{A}^c) \leq \frac{(2e \log R - e)C}{(\log R)^{1+\delta}}. \quad \square$$

Proof of Theorem 3.3.3 The proofs of this theorem are similar to the one of Theorem 3.3.1 up to a modification of the bound on $P(\mathcal{A}^c)$ in Lemma 3.3.1. We consider now the event

$$\mathcal{A} = \left\{ \max_{1 \leq r \leq R} \left\{ \sum_{t=1}^T \sum_{k=1}^K \left(\sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\}^{1/2} \leq \lambda JT \right\}.$$

Thus, following the fact that different space basis Ψ_k and $\Psi_{k'}$ are independent, we have:

$$\begin{aligned}
P(\mathcal{A}^c) &= P \left[\max_{1 \leq r \leq R} \left\{ \sum_{t=1}^T \sum_{k=1}^K \left(\sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\}^{1/2} > \lambda J T \right] \\
&= P \left[\max_{1 \leq r \leq R} \left\{ \sum_{t=1}^T \left(\sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\}^{1/2} > \lambda J T \right] \\
&\leq R P \left[\left\{ \sum_{t=1}^T \left(\sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\}^{1/2} > \lambda J T \right] \\
&= R P \left[\left\{ \sum_{t=1}^T \left(\sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \right)^2 \right\}^{1/2} > \lambda J T \right] \\
&= R P \{ f(\underline{V}) > \lambda \}
\end{aligned}$$

where

$$\begin{aligned}
V_t &\stackrel{\text{def}}{=} J^{-1} \sum_{k=1}^K \sum_{j=1}^J \Psi_{tkj} \varepsilon_{tj} U_{tr} \\
\underline{V} &\stackrel{\text{def}}{=} (V_{1r}, \dots, V_{Tr}) \\
f(\underline{V}) &\stackrel{\text{def}}{=} T^{-1} \left(\sum_{t=1}^T V_t^2 \right)^{1/2}.
\end{aligned}$$

Since Assumption 3.3.3 holds, i.e. with a high probability p , for $\forall t$ and $v_{1r}, \dots, v_{Tr}, v'_{tr}$,

$$\begin{aligned}
|f(v_{1r}, \dots, v_{tr}, \dots, v_{Tr}) - f(v_{1r}, \dots, v'_{tr}, \dots, v_{Tr})| &\leq b_t^2 / T \\
\mathbb{E} f(\underline{V}) &\leq \frac{C'}{\sqrt{T}}.
\end{aligned}$$

Then, by the (extended) Mcdiarmid inequality, see Theorem 2.1 of Janson (2004), with the random vectors \underline{V} and function f , we have

$$\begin{aligned}
P(\mathcal{A}^c) &\leq R P \{ f(\underline{V}) > \lambda \} \leq R P \{ f(\underline{V}) - \mathbb{E} f(\underline{V}) > \lambda - \frac{C'}{\sqrt{T}} \} \\
&\leq R \exp \left\{ - \frac{(\lambda - \frac{C'}{\sqrt{T}})^2 T^2}{\mathcal{X}^*(\mathcal{T}) \sum_t b_t^2} \right\} = R^{-\delta'}
\end{aligned}$$

$$\text{with } \lambda = \frac{C'}{\sqrt{T}} + \sqrt{\frac{\mathcal{X}^*(\mathcal{T}) \sum_t b_t^2}{(\log R)^{1-\delta'} T^2}}, \quad \delta' > 0. \quad \square$$

3 Generalized Dynamic Semiparametric Factor Models

Proof of Theorem 3.4.1 Similar to $\hat{Y}_t^\top \stackrel{\text{def}}{=} Y_t^\top - U_t^\top \hat{\beta} \Psi_t$, define $\tilde{Y}_t^\top \stackrel{\text{def}}{=} Y_t^\top - U_t^\top \beta^* \Psi_t$ with the corresponding estimate $\tilde{Z}_{0,t}$. Thus

$$\frac{1}{T} \sum_{1 \leq t \leq T} \left\| \hat{Z}_{0,t}^\top \hat{A} - Z_{0,t}^\top A^* \right\|^2 \leq \frac{1}{T} \sum_{1 \leq t \leq T} \left\| \hat{Z}_{0,t}^\top \hat{A} - \tilde{Z}_{0,t}^\top \hat{A} \right\|^2 + \frac{1}{T} \sum_{1 \leq t \leq T} \left\| \tilde{Z}_{0,t}^\top \hat{A} - Z_{0,t}^\top A^* \right\|^2,$$

where the second term is bounded by $\mathcal{O}_P(\rho^2 + \delta_K^2)$ by Theorem 2 of Park et al. (2009). For the first term, since

$$\begin{aligned} \hat{Z}_{0,t} &= (\hat{A} \Psi_t \Psi_t^\top \hat{A}^\top)^{-1} \hat{A} \Psi_t \hat{Y}_t \\ \tilde{Z}_{0,t} &= (\hat{A} \Psi_t \Psi_t^\top \hat{A}^\top)^{-1} \hat{A} \Psi_t \tilde{Y}_t \\ \tilde{Z}_{0,t} - \hat{Z}_{0,t} &= (\hat{A} \Psi_t \Psi_t^\top \hat{A}^\top)^{-1} \hat{A} \Psi_t \{ \Psi_t^\top (\hat{\beta} - \beta^*) U_t \} \end{aligned}$$

and Theorem 3.3.3 tells us that $(JT)^{-1} \sum_{t=1}^T \left\| \Psi_t^\top (\hat{\beta} - \beta^*) U_t \right\|^2$ could be arbitrary small, i.e. \exists large enough R , s.t. the first term is dominated by the second one. \square

Proof of Theorem 3.4.2 The proof is in a similar spirit of the one of Theorem 3 in Park et al. (2009). We will prove the first equation of the theorem for $h \neq 0$. The second equation follows from the first equation. We first prove that the matrix $T^{-1} \sum_{t=1}^T Z_{0,t} \hat{Z}_{0,t}^\top$ is invertible. Suppose that the assertion is not true. We can choose a random vector e such that $\|e\| = 1$ and $e^\top \sum_{t=1}^T Z_{0,t} \hat{Z}_{0,t}^\top = 0$. Note that

$$\begin{aligned} & \left\| T^{-1} \sum_{t=1}^T Z_{0,t} \hat{Z}_{0,t}^\top \hat{A} - T^{-1} \sum_{t=1}^T Z_{0,t} Z_{0,t}^\top A^* \right\| \\ & \leq T^{-1} \sum_{t=1}^T \| Z_{0,t} (\hat{Z}_{0,t}^\top \hat{A} - Z_{0,t}^\top A^*) \| \\ & \leq (T^{-1} \sum_{t=1}^T \| Z_{0,t} \|^2)^{1/2} (T^{-1} \sum_{t=1}^T \|\hat{Z}_{0,t}^\top \hat{A} - Z_{0,t}^\top A^*\|^2)^{1/2} \\ & = \mathcal{O}_P(\rho + \delta_K), \end{aligned} \tag{3.28}$$

because of Assumption (4.1.5) and Theorem 3.4.1. Thus with $f = T^{-1} \sum_{t=1}^T Z_{0,t} Z_{0,t}^\top e$, we obtain

$$\begin{aligned} \|f^\top m\| &= \|f^\top (A^* \Psi)\| + \mathcal{O}_P(\delta_K) \\ &= \|e^\top T^{-1} \sum_{t=1}^T Z_{0,t} Z_t^\top \hat{A} \Psi\| + \mathcal{O}_P(\rho + \delta_K) \\ &= \mathcal{O}_P(\rho + \delta_K). \end{aligned}$$

This implies that m_1, \dots, m_L are linearly dependent, contradicting to the construction that all space basis are independent.

$\tilde{Z}_{0,t} = B^\top \hat{Z}_{0,t}$ and $\tilde{A} = B^{-1}A$ give with (3.28)

$$\begin{aligned}
\|\tilde{A} - A^*\| &= \|T^{-1} \sum_{t=1}^T Z_{0,t} Z_t^\top (\tilde{A} - A^*)\| \mathcal{O}_P(1) \\
&= \|T^{-1} \sum_{t=1}^T Z_{0,t} \tilde{Z}_{0,t}^\top \tilde{A} - T^{-1} \sum_{t=1}^T Z_{0,t} Z_{0,t}^\top A^*\| \mathcal{O}_P(1) \\
&= \mathcal{O}_P(\rho + \delta_K)
\end{aligned} \tag{3.29}$$

From Assumptions (4.1.4), (3.29) and Theorem 3.4.1, we get

$$\begin{aligned}
&T^{-1} \sum_{t=1}^T \|\tilde{Z}_t^\top - Z_{0,t}\|^2 \\
&= T^{-1} \sum_{t=1}^T \|\tilde{Z}_t^\top (m_1, \dots, m_L)^\top - Z_{0,t}^\top (m_1, \dots, m_L)^\top\|^2 \mathcal{O}_P(1) \\
&= T^{-1} \sum_{t=1}^T \|\tilde{Z}_t^\top A^* - \tilde{Z}_t^\top \tilde{A}\|^2 \mathcal{O}_P(1) \\
&\quad + T^{-1} \sum_{t=1}^T \|\tilde{Z}_t^\top \tilde{A} - Z_{0,t}^\top A^*\|^2 \mathcal{O}_P(1) + \mathcal{O}_P(\delta_K^2) \\
&\leq T^{-1} \sum_{t=1}^T \|\tilde{Z}_{0,t} - Z_{0,t}\|^2 \|\tilde{A} - A^*\|^2 \mathcal{O}_P(1) \\
&\quad + T^{-1} \sum_{t=1}^T \|Z_{0,t}\|^2 \|\tilde{A} - A^*\|^2 \mathcal{O}_P(1) \\
&\quad + T^{-1} \sum_{t=1}^T \|\tilde{Z}_t^\top \tilde{A} - Z_{0,t}^\top A^*\|^2 \mathcal{O}_P(1) + \mathcal{O}_P(\delta_K^2) \\
&= \mathcal{O}_P(\rho^2 + \delta_K^2).
\end{aligned} \tag{3.30}$$

We will show that for $h \neq 0$

$$T^{-1} \sum_{t=h+1}^T \{(\tilde{Z}_{0,t+h} - Z_{0,t+h}) - (\tilde{Z}_{0,t} - Z_{0,t})\} Z_{0,t}^\top = \mathcal{O}_P(T^{-1/2}) \tag{3.31}$$

This implies the first statement of Theorem 3.4.2, because by (3.30)

$$T^{-1} \sum_{t=-h+1}^T (\tilde{Z}_{0,t} - Z_{0,t})(\tilde{Z}_{0,t+h} - Z_{0,t+h}) = \mathcal{O}_P(b^2) = \mathcal{O}_P(T^{-1/2}).$$

3 Generalized Dynamic Semiparametric Factor Models

For the proof of (3.31), define

$$\begin{aligned}
\tilde{S}_{t,Z} &= J^{-1} \sum_{j=1}^J \tilde{A} \Psi(X_{t,j}) \Psi(X_{t,j})^\top \tilde{A}^\top \\
S_{t,Z} &= A^* E \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top \right\} A^{*\top} \\
\tilde{S}_\alpha &= (JT)^{-1} \sum_{t=1}^T \sum_{j=1}^J \{ \Psi(X_{t,j}) \otimes \tilde{Z}_{0,t} \} \{ \Psi(X_{t,j}) \otimes \tilde{Z}_{0,t} \}^\top \\
S_\alpha &= T^{-1} \sum_{t=1}^T E \left[\{ \Psi(X_{t,j}) \otimes Z_{0,t} \} \{ \Psi(X_{t,j}) \otimes Z_{0,t} \}^\top | Z_{0,t} \right] \\
S &= J^{-1} A^* \left[\Psi(X_{t,j}) \Psi(X_{t,j})^\top e - E \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top e \right\} \right],
\end{aligned}$$

where $e \in \mathbb{R}^K$ with $\|e\| = 1$. Let \tilde{a} be the stack form of \tilde{A} . It can be verified that

$$\tilde{Z}_{0,t} = \tilde{S}_{t,Z}^{-1} J^{-1} \sum_{j=1}^J \{ Y_{t,j} A \Psi(X_{t,j}) \}, \quad (3.32)$$

$$\tilde{a} = \tilde{S}_\alpha^{-1} (JT)^{-1} \sum_{t=1}^T \sum_{j=1}^J \{ \Psi(X_{t,j}) \otimes \tilde{Z}_{0,t} \} Y_{t,j}. \quad (3.33)$$

Let $\gamma = T^{-1/2}/b$. We argue that

$$\sup_{1 \leq t \leq T} \|\tilde{S}_{t,Z} - S_{t,Z}\| = \mathcal{O}_P(\gamma), \quad \|\tilde{S}_\alpha - S_\alpha\| = \mathcal{O}_P(\gamma). \quad (3.34)$$

We show the first part of (3.34). The second part can be shown similarly. Since

$$\tilde{A} \Psi_t \Psi_t^\top \tilde{A}^\top = (\tilde{A} - A^* + A^*) (\Psi_t \Psi_t^\top - E \Psi_t \Psi_t^\top + E \Psi_t \Psi_t^\top) (\tilde{A} - A^* + A^*)^\top,$$

to prove the first part it suffices to show that, uniformly for $1 \leq t \leq T$,

$$J^{-1} \sum_{j=1}^J A^* \left[\Psi(X_{t,j}) \Psi(X_{t,j})^\top - \mathbb{E} \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top \right\} \right] (\tilde{A} - A^*)^\top = \mathcal{O}_P(\gamma) \quad (3.35)$$

$$J^{-1} \sum_{j=1}^J (\tilde{A} - A^*) \left[\Psi(X_{t,j}) \Psi(X_{t,j})^\top - \mathbb{E} \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top \right\} \right] (\tilde{A} - A^*)^\top = \mathcal{O}_P(\gamma) \quad (3.36)$$

$$J^{-1} \sum_{j=1}^J A^* \left[\Psi(X_{t,j}) \Psi(X_{t,j})^\top - \mathbb{E} \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top \right\} \right] A^{*\top} = \mathcal{O}_P(\gamma) \quad (3.37)$$

$$J^{-1} \sum_{j=1}^J A^* \mathbb{E} \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top \right\} (\tilde{A} - A^*)^\top = \mathcal{O}_P(\gamma) \quad (3.38)$$

$$J^{-1} \sum_{j=1}^J (\tilde{A} - A^*) \mathbb{E} \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top \right\} (\tilde{A} - A^*)^\top = \mathcal{O}_P(\gamma) \quad (3.39)$$

The proof of (3.35)-(3.37) follows by simple arguments. We now show (3.38). Claim (3.39) can be shown similarly. For the proof of (3.38), we use Bernstein's inequality for the following sum:

$$\mathbb{P} \left(\left| \sum_{j=1}^J W_j \right| > x \right) \leq 2 \exp \left(-\frac{1}{2} \frac{x^2}{V} + Mx/3 \right). \quad (3.40)$$

Here for a value of t with $1 \leq t \leq T$, the random variable W_j is an element of the $L \times 1$ -matrix $S = J^{-1} A^* \left[\Psi(X_{t,j}) \Psi(X_{t,j})^\top e - \mathbb{E} \left\{ \Psi(X_{t,j}) \Psi(X_{t,j})^\top e \right\} \right]$ where $e \in \mathbb{R}^K$ with $\|e\| = 1$. In (3.40), V is an upper bound for the variance of $\sum_{j=1}^J W_j$ and M is a bound for the absolute values of W_j , i.e. $|W_j| \leq M$ for $1 \leq j \leq J$, a.s. With some constants C_1 and C_2 that do not depend on t and the row number we get $V \leq C_1 J^{-1}$ and $M \leq C_2 K^{1/2} J^{-1}$. Application of Bernstein's inequality gives that, uniformly for $1 \leq t \leq T$ and $e \in \mathbb{R}^K$ with $\|e\| = 1$, all L elements of S are of order $\mathcal{O}_P(\gamma)$. This shows claim (3.35).

From (3.29), (3.30), (3.32), (3.33) and (3.34) it follows that uniformly for $1 \leq t \leq T$,

$$\begin{aligned} \tilde{Z}_{0,t} - Z_{0,t} &= S_{t,Z}^{-1} J^{-1} \sum_{j=1}^J \varepsilon'_{t,j} A^* \Psi(X_{t,j}) \\ &\quad + S_{t,Z}^{-1} J^{-1} \sum_{j=1}^J \varepsilon'_{t,j} (\tilde{A} - A^*) \Psi(X_{t,j}) + \mathcal{O}_P(T^{-1/2}) \\ &\stackrel{\text{def}}{=} \Delta_{t,1,Z} + \Delta_{t,2,Z} + \mathcal{O}_P(T^{-1/2}). \end{aligned} \quad (3.41)$$

3 Generalized Dynamic Semiparametric Factor Models

For the proof of the theorem it remains to show that for $1 \leq j \leq 2$

$$T^{-1} \sum_{t=-h+1}^T (\Delta_{t+h,j,Z} - \Delta_{t,j,Z}) Z_{0,t}^\top = \mathcal{O}_P(T^{-1/2}). \quad (3.42)$$

This can be easily checked for $j = 1$. For $j = 2$ it follows from $\|\tilde{A} - A^*\| = \mathcal{O}_P(\rho + \delta_K)$ and

$$\mathbb{E} \left\{ \left\| (JT)^{-1} \sum_{t=1}^T \sum_{j=1}^J \varepsilon'_{t,j} S_{t,Z}^{-1} M \Psi(X_{t,j}) \right\|^2 \right\} = \mathcal{O}(K(JT)^{-1}),$$

for any $L \times K$ matrix M with $\|M\| = 1$. \square

Bibliography

- G. Becker. Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, 3rd edition. The University of Chicago Press, 1994.
- A. Belloni and V. Chernozhukov. l_1 -penalized quantile regression in high-dimensional sparse models. CEMMAP Working Paper, 10/09, 2009.
- F. Benth and J. Benth. Stochastic modelling of temperature variations with a view towards weather derivatives. Applied Mathematical Finance, 12(1):53–85, 2005.
- J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. Ann. Statist., 36(1):199–227, 2008a.
- J. Bickel and Elizaveta Levina. Covariance regularization by thresholding. Ann. Statist., 36(6):2577–2604, 2008b.
- P. Bickel and M. Rosenblatt. On some global measures of the deviation of density function estimators. Annals of Statistics, 1:1071–1095, 1973.
- P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. Annals of Statistics, to appear, 2008.
- F. Black and M. Scholes. The pricing of options and corporate liabilities. Journal of Political Economy, 81:637–654, 1973.
- S. Borak, W. Härdle, and B. Lopez. Statistics of Financial Markets Exercises and Solutions. Springer-Verlag, Heidelberg, 2010.
- M. Buchinsky. Quantile regression, box-cox transformation model, and the u.s. wage structure, 1963–1987. Journal of Econometrics, 65:109–154, 1995.
- Z. W. Cai. Regression quantiles for time series. Econometric Theory, 18:169–192, 2002.
- V. Chernozhukov, S. Lee, and A. M. Rosen. Intersection bounds: estimation and inference. CEMMAP Working Paper, 19/09, 2009.
- S. Csörgö and P. Hall. Upper and lower classes for triangular arrays. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 61:207–222, 1982.
- J. C. Day and E. C. Newburger. The big payoff: Educational attainment and synthetic estimates of work-life earnings. special studies. current population reports. Statistical report p23-210, U.S. Department of Commerce Economics and Statistics Administration, U.S. CENSUS BUREAU, 2002.

Bibliography

- L. Denby. Smooth regression functions. Statistical report 26, AT&T Bell Laboratories, 1986.
- F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. Journal of Econometrics, 130:337–364, 2006.
- L. Dümbgen, S. van de Geer, M. Veraar, and J. A. Wellner. Nemirovski’s Inequalities Revisited. ArXiv e-prints, July 2008.
- C. Dustmann, J. Ludsteck, and U. Schönberg. Revisiting the German wage structure. Quarterly Journal of Economics, forthcoming, 2009.
- J. Fan, T. C. Hu, and Y. K. Truong. Robust nonparametric function estimation. Scandinavian Journal of Statistics, 21:433–446, 1994.
- J. Fan, Q. Yao, and H. Tong. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika, 83:189–206, 1996.
- M. R. Fengler, W. Härdle, and E. Mammen. A semiparametric factor model for implied volatility surface dynamics. Journal of Financial Econometrics, 5(2):189–218, 2007.
- T. S. Ferguson. Mathematical Statistics: A Decision Theoretic Approach. New York: Academic Press, 1967.
- R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Cambridge Philos. Soc., 24:180–190, 1928.
- Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. Journal of the American Statistical Association, 100:830–840, September 2005. URL <http://ideas.repec.org/a/bes/jnlasa/v100y2005p830-840.html>.
- J. Franke and P. Mwita. Nonparametric estimates for conditional quantiles of time series. Report in Wirtschaftsmathematik 87, University of Kaiserslautern, 2003.
- Wenjiang J. Fu. Penalized regressions: The bridge versus the Lasso. Journal of Computational and Graphical Statistics, 7(3):397–416, 1998. ISSN 10618600. URL <http://www.jstor.org/stable/1390712>.
- T. Gasser, R. Möcks, and R. Verleger. Selavco: A method to deal with trial-to-trial variability of evoked potential. Electroencephalography and Clinical Neurophysiology, 55:717–723, 1983.
- Domenico Giannone, Lucrezia Reichlin, and Luca Sala. Monetary policy in real time. In NBER Macroeconomics Annual 2004, Volume 19, NBER Chapters, pages 161–224. National Bureau of Economic Research, Inc, 2005. URL <http://ideas.repec.org/h/nbr/nberch/6670.html>.

- P. H. Gleick et al. Climate change and the integrity of science. Science, 328:689–691, 2010.
- P. J. Green and B. S. Yandell. Semi-Parametric Generalized Linear Models, volume 32 of Lecture Notes in Statistics 32. Springer, New York, 1985.
- Jinyong Hahn. Bootstrapping quantile regression estimators. Econometric Theory, 11(1):105–121, 1995.
- Anthony Hall and Nikolaus Hautsch. Order aggressiveness and order book dynamics. Empirical Economics, 30(4):973–1005, January 2006.
- P. Hall. On convergence rates of suprema. Probab. Th. Rel. Fields, 89:447–455, 1991.
- P. Hall, R.C.L. Wolff, and Q. Yao. Methods for estimating a conditional distribution function. Journal of the American Statistical Association, 94:154–163, 1999.
- P. Hall, H. G. Müller, and J. L. Wang. Properties of principal component methods for functional and longitudinal data analysis. The Annals of Statistics, 34(3):1493–1517, 2006.
- W. Härdle. Asymptotic maximal deviation of M-smoothers. Journal of Multivariate Analysis, 29:163–179, 1989.
- W. Härdle and S. Luckhaus. Uniform consistency of a class of regression function estimators. Annals of Statistics, 12:612–623, 1984.
- W. Härdle and J. Marron. Bootstrap simultaneous error bars for nonparametric regression. Ann. Statist., 19:778–796, 1991.
- W. Härdle and S. Song. Confidence bands in quantile regression. Econometric Theory, 26:1–22, 2010.
- W. Härdle, P. Janssen, and R. Serfling. Strong uniform consistency rates for estimators of conditional functionals. Annals of Statistics, 16:1428–1429, 1988.
- Nikolaus Hautsch and Yangguoyi Ou. Yield curve factors, term structure volatility, and bond risk premia. SFB 649 Discussion Papers SFB649DP2008-053, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, July 2008. URL <http://ideas.repec.org/p/hum/wpaper/sfb649dp2008-053.html>.
- A. E. Hedin. Extension of the msis thermosphere model into the middle and lower atmosphere. Journal of Geophysical Research, 96:1159–1172, feb 1991.
- Joel L. Horowitz. Bootstrap methods for median regression models. Econometrica, 66(6):1327–1351, 1998.
- P. Huber. Robust Statistics. Wiley, New York., 1981.

Bibliography

- S. Janson. Large deviations for sums of partly dependent random variables. Random Structures Algorithms, 24(3):234–248, 2004.
- Kiho Jeong, W. Härdle, and S. Song. A consistent nonparametric test for causality in quantile. Econometric Theory, revise and resubmit, 2009.
- J. Jia, K. Rohe, and B. Yu. The lasso under heteroscedecity. Technical Report 783, Statistics Department, UC Berkeley, 2009.
- G. Johnston. Probabilities of maximal deviations of nonparametric regression function estimates. Journal of Multivariate Analysis, 12:402–414, 1982.
- T. R. Karl, G. Kukla, V. N. Razuvayev, M. J. Changery, R. G. Quayle, R. R. Heim, and D. R. Easterling. Global warming: Evidence for asymmetric diurnal temperature change. Geophysical Research Letters, 18:2253–2256, December 1991. doi: 10.1029/91GL02900.
- G. Kauermann. Modeling longitudinal data with ordinal response by varying coefficients. Biometrics, 56(3):1692–698, 2000.
- K. Knight. Comparing conditional quantile estimators: first and second order considerations. Unpublished manuscript, 2001.
- R. Koenker and G. W. Bassett. Regression quantiles. Econometrica, 46:33–50, 1978.
- R. Koenker and K. F. Hallock. Quantile regression. Journal of Econometric Perspectives, 15(4):143–156, 2001.
- R. Koenker and B. J. Park. An interior point algorithm for nonlinear quantile regression. Journal of Econometrics, 71:265–283, 1996.
- E. Kong, O. Linton, and Y. Xia. Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. Econometric Theory, accepted, 2008.
- C. Kuan, J. Yeh, and Y. Hsu. Assessing value at risk with care, the conditional autoregressive expectile models. Journal of Econometrics, 150:261–270, 2009.
- R. D. Lee and L. Carter. Modeling and forecasting the time series of u.s. mortality. Journal of the American Statistical Association, 87(419):659–671, 1992.
- M. G. Lejeune and P. Sarda. Quantile regression: A nonparametric approach. Computational Statistics and Data Analysis, 6:229–239, 1988.
- H. Liang and R. Li. Variable selection for partially linear models with measurement errors. J. Amer. Statist. Assoc., forthcoming, 2009.
- K. Lounici. Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. Electronic Journal of Statistics, 2:90–102, 2008.

- Karim Lounici, Massimiliano Pontil, Alexandre B. Tsybakov, and Sara van de Geer. Taking advantage of sparsity in multi-task learning. Proceedings of Conference on Learning Theory (COLT) 2009, 2009. URL <http://www.citebase.org/abstract?id=oai:arXiv.org:0903.1468>.
- T. Martinussen and T. Scheike. A nonparametric dynamic additive regression model for longitudinal data. Annals of Statistics, 28(4):1000–1025, 2000.
- Peter N. C. Mohr, Guido Biele, Lea K. Krugel, Shu-Chen Li, and Hauke R. Heekeren. Neural foundations of risk-return trade-off in investment decisions. NeuroImage, 49(3):2556–2563, February 2010. URL <http://dx.doi.org/10.1016/j.neuroimage.2009.10.060>.
- Kevin Murphy and Finis Welch. Empirical age-earnings profiles. Journal of Labor Economics, 8(2):202–229, 1990.
- Alena Myšičková, Song Song, Peter N.C. Mohr, Hauke R. Heekeren, and Wolfgang K. Härdle. Risk patterns and correlated brain activities. Submitted to Neuroimage, 2010.
- C. R. Nelson and A. F. Siegel. Parsimonious modeling of yield curves. Journal of Business, 60:473–489, 1987.
- W. Newey and J. Powell. Asymmetric least squares estimation and testing. Econometrica, 55:816–847, 1987.
- M. Odening, E. Berg, and C.G. Turvey. Management of climate risk in agriculture. Special Issue of the Agricultural Finance Review, 68(1):83–97, 2008.
- Byeong U. Park, Enno Mammen, Wolfgang Härdle, and Szymon Borak. Time series modelling with semiparametric factor dynamics. Journal of the American Statistical Association, 104(485):284–298, 2009. URL <http://econpapers.repec.org/RePEc:bes:jnlasa:v:104:i:485:y:2009:p:284-298>.
- William J. Parton and Jesse A. Logan. A model for diurnal variation in soil and air temperature. Agricultural Meteorology, 23:205 – 216, 1981. ISSN 0002-1571. doi: DOI:10.1016/0002-1571(81)90105-9. URL <http://www.sciencedirect.com/science/article/B757B-48BC9WN-P/2/b6c8cacf7f00f2a6639ff85cc3fb3d80>.
- M. Parzen. On estimation of a probability density function and mode. Annals of Mathematical Statistics, 32:1065–1076, 1962.
- D. Pollard. Asymptotics for least absolute deviation estimators. Econometric Theory, 7:186–199, 1991.
- S. Portnoy and R. Koenker. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute -error estimators (with discussion). Statistical Sciences, 12:279–300, 1997.

Bibliography

- P. Racsco, L. Szeidl, and M. Semenov. A serial approach to local stochastic weather models. Ecological Modelling, 57(1-2):27 – 41, 1991. ISSN 0304-3800. doi: DOI:10.1016/0304-3800(91)90053-4. URL <http://www.sciencedirect.com/science/article/B6VBS-48XDC36-56/2/7f98e0ac150cbd45ffb9c17f95de8550>.
- J. Ramway and B. Silverman. Functional Data Analysis, 2nd Edition. Springer, 2005.
- P. M. Robinson. Semiparametric econometrics: A survey. Journal of Applied Econometrics, 3:35–51, 1988.
- M. Rosenblatt. Remarks on a multivariate transformation. Annals of Mathematical Statistics, 23:470–472, 1952.
- D. Ruppert, S. J. Sheather, and M. P. Wand. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association, 90:1257–1270, 1995.
- P. E. Speckman. Regression analysis for partially linear models. Journal of the Royal Statistical Society, Series B, 50:413–436, 1988.
- J. H. Stock and M. W. Watson. Implications of dynamic factor models for var analysis. NBER Working Papers 11467, National Bureau of Economic Research, Inc, July 2005. available at <http://ideas.repec.org/p/nbr/nberwo/11467.html>.
- James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association, 97:1167–1179, December 2002a. URL <http://ideas.repec.org/a/bes/jnlasa/v97y2002mdecemberp1167-1179.html>.
- James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics, 20(2):147–62, April 2002b. URL <http://ideas.repec.org/a/bes/jnlbes/v20y2002i2p147-62.html>.
- Robert Tibshirani. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B, 58(1):267–288, 1996. ISSN 00359246. URL <http://www.jstor.org/stable/2346178>.
- G. Tusnady. A remark on the approximation of the sample distribution function in the multidimensional case. Periodica Mathematica Hungarica, 8:53–55, 1977.
- K.J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A general statistical analysis for fmri data. NeuroImage, 15:1–15, 2002.
- K. Yu and M. C. Jones. A comparison of local constant and local linear regression quantile estimation. Computational Statistics and Data Analysis, 25:159–166, 1997.
- K. Yu and M. C. Jones. Local linear quantile regression. Journal of the American Statistical Association, 93:228–237, 1998.

- K. Yu, Z. Lu, and J. Stander. Quantile regression: applications and current research areas. Journal of the Royal Statistical Society: Series D, 52:331–350, 2003.
- Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B, 68(1):49–67, 2006. URL <http://ideas.repec.org/a/bla/jorssb/v68y2006i1p49-67.html>.

List of Figures

1.1	The 0.5-quantile curve, the NW estimator $m_n^*(x)$, and the 0.5-quantile smoother $l_n(x)$	9
1.2	The 0.9-quantile curve, the 0.9-quantile smoother and 95% confidence band.	10
1.3	Original observations, local quantiles, quantile smoothers and corresponding confidence bands.	11
1.4	Quadratic, quartic, set of dummies estimates, quantile smoothers and confidence bands.	12
2.1	Real quantile curve, quantile estimate with asymptotic and bootstrapping confidence band.	30
2.2	$\hat{\beta}$ with respect to different quantiles for different numbers of observations.	32
2.3	Nonparametric part smoothing, real curve, smoother with bootstrap confidence band.	33
2.4	Boxplots for “low education”, “apprenticeship” & “university” groups.	35
2.5	$\hat{\beta}$ corresponding to different quantiles with 6, 13, 25 partitions.	36
2.6	95% uniform confidence bands for 0.05-quantile smoothers with 3 different education levels	37
2.7	95% uniform confidence bands for 0.50-quantile smoothers with 3 different education levels	37
2.8	95% uniform confidence bands for 0.99-quantile smoothers with 3 different education levels	38
3.1	Temperatures of China and the corresponding moving average (of 730 nearby days) view (right).	48
3.2	Typical fMRI data in one particular time point.	48
3.3	The typical IV data design on two different days.	49
3.4	Space basis using the series estimator for the IVS modeling.	50
3.5	An illustration plot about how group Lasso penalty shrinks the coefficients.	60
3.6	Distribution of the eigenvalues and the relative proportion of explained variance.	62
3.7	China Climate Types	62
3.8	Estimated coefficients of the $54 \cdot 3$ yearly polynomial time basis.	64
3.9	Extracted trends based on $U_t^\top \hat{\beta}$	65
3.10	Estimated Stochastic Process $\hat{Z}_{0,t,1}$ and a 10 year zoom.	65
3.11	Flowchart of the experiment (left) and risk attitudes.	67
3.12	Estimated function \hat{m}_3 with highlighted VMPFC.	68
3.13	Response curves (to stimuli) $U_t^\top \hat{\Gamma}_2^i$	69

List of Figures

3.14 Space basis and the estimated time series of factors $\hat{Z}_{t,1}, \hat{Z}_{t,2}$	70
--	----

List of Tables

2.1	Simulated coverage probabilities & areas of confidence bands.	30
2.2	SSE of $\hat{\beta}$ with respect to a_n for different numbers of observations.	31
3.1	L_{par} and L_{pre} w.r.t. different sparsity levels.	60
3.2	L_{par} and L_{pre} w.r.t. different number of initially involved time basis. . . .	61
3.3	L_{par} and L_{pre} w.r.t. different levels of dependence of ε_t	61
3.4	Initial choice of $53 \cdot 3 + 20 = 179$ time basis.	62
3.5	Estimated coefficients of the 5 factors w.r.t. the 20 Fourier series time basis.	63
3.6	Classification rates of the SVM method.	69

Declaration of Authorship

I hereby confirm that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works of any other author, in any form, is properly acknowledged at their point of use.

Berlin, den 28.01.2008

Song Song